REPORT RESUMES

ED 017 015 24 CG 001 685

THE MEASUREMENT AND EFFECT OF RISK TAKING ON OBJECTIVE EXAMINATIONS. FINAL REPORT.

BY- SLAKTER, MALCOLM J.
STATE UNIV. OF N.Y., BUFFALO

REPORT NUMBER BR-5-8428 PUB DATE SEP 67
CONTRACT OEC-6-10-239
EDRS PRICE MF-$0.50 HC-$4.08 100P.

DESCRIPTORS- *RISK, *STUDENT BEHAVIOR, TEST CONSTRUCTION, *OBJECTIVE TESTS, *MEASUREMENT TECHNIQUES, *TEST VALIDITY, TEST RELIABILITY, RESEARCH DESIGN, RESEARCH METHODOLOGY,

RISK TAKING ON OBJECTIVE EXAMINATIONS (RTOOE) OCCURS WHENEVER AN EXAMINEE TAKES AN EXAMINATION WITH PENALTIES FOR INCORRECT RESPONSES, AND RESPONDS TO AN ITEM WITHOUT KNOWING THE CORRECT ANSWER WITH CERTAINTY. THIS PROJECT WAS CONCERNED WITH THREE ASPECTS OF RTOOE--(1) ITS MEASUREMENT, (2) ITS GENERALITY, AND (3) ITS EFFECT ON TEST SCORE. IN THE STUDY OF THE MEASUREMENT OF RTOOE, A NEW INDEX WAS INTRODUCED, AND ITS CONCURRENT VALIDITY INVESTIGATED BY EXAMINING THE CORRELATION OF THE NEW INDEX WITH THE ZILLER INDEX. MUCH OF THE RELIABLE VARIANCE OF THE ZILLER INDEX WAS PREDICTABLE FROM THE NEW INDEX, AND BOTH WERE ESSENTIALLY UNCORRELATED WITH LEGITIMATE SCORE. IT WAS CONCLUDED THAT THE NEW INDEX WAS A POTENTIALLY VALUABLE MEASURE OF RTOOE. THE GENERALITY OF RTOOE WAS EXAMINED ACROSS FOUR TYPES OF TESTS. EACH S RECEIVED TWO RTOOE SCORES (NEW AND ZILLER), AND ONE LEGITIMATE SCORE ON EACH OF THE FOUR TESTS. THE RESULTING CORRELATION MATRIX INDICATED THAT THERE WAS A STRONG GENERAL RTOOE FACTOR. TO INVESTIGATE THE EFFECT OF RTOOE ON TEST SCORE, RTOOE WAS MANIPULATED, AND THE EFFECT ON AVERAGE TEST SCORE NOTED. IN EACH CASE, THE RTOOE EFFECTS WERE SIGNIFICANT AT THE .05 LEVEL, WITH THE MORE CAUTIOUS RTOOE STRATEGIES RESULTING IN THE LOWER MEAN TEST SCORES. (AUTHOR)

ED017015

FINAL REPORT
Project No. 5-8428 - 24
Contract No. OE-6-10-239

THE MEASUREMENT AND EFFECT OF RISK TAKING ON OBJECTIVE EXAMINATIONS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

September 1967

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

CG 001 685

The Measurement and Effect of Risk Taking on Objective Examinations

Project No. 5-8428
Contract No. OE-6-10-239

Malcolm J. Slakter

September 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

State University of New York at Buffalo

Buffalo, New York

Acknowledgments

Research in the public schools generally requires the cooperation of a large set of people. This project was certainly no exception, and the author is indebted to many people, but especially the following:

1. The hundreds of students who served as subjects for the study.

2. Messrs James Allan, Edward Allen, Samuel Bennett, Joseph Castiglia, Stewart Davis, Roger Farr, William Ferrara, G. J. Goman, Gerald Gougeon, Harold Johnson, William Keller, James Laas, Len Popp, Oscar Roaldi, William Sawin, D. J. Schrader, Stuart Tuck, Fred Tuthill, Richard Warner, Daniel Weppner, and Miss Helen Kennedy, for their assistance in obtaining the subjects, and in the test administration.

3. Carl Terranova for his sizable contributions in the construction of the examinations.

4. Gail Hofmann for her precise typing of the manuscript, and her usual competence in all related matters.

5. Roger Koehler for his invaluable assistance in all stages of the project: test construction, test administration, computer programming in the analysis of the data, numerous comments and suggestions on the final report, proof reading of the final report.

September, 1967 M.J.S.

CONTENTS

Chapter I Introduction 2

Chapter II Measurement of RTOOE 17

Chapter III Generality of RTOOE 24

Chapter IV Effect of RTOOE on Test Score 33

Chapter V Summary 41

References 44

Appendixes

A Directions for Administering, Standard Educational Intelligence Achievement Test, Grades 7-8-9 47

B Standard Educational Intelligence Test, Grades 7-8-9 50

C Standard Educational Achievement Test, Grades 7-8-9 62

D Directions for Administering, Standard Educational Intelligence Test, Grades 10-11-12 75

E Directions to Subjects, Standard Educational Intelligence Test, Grades 10-11-12 78

F Test Items, Standard Educational Intelligence Test, Grades 10-11-12 92

Tables

2.1 Means and Standard Deviations for Vocabulary and Risk Taking Measures 20

2.2 Split-half Reliabilities for Vocabulary and Risk Taking Measures 22

2.3 Correlations for the Vocabulary and Risk Measures 23

3.1 School Means on Risk (R', R_z) and Legitimate Measures (L) 28

Tables

3.2 Correlations and Reliabilities for Risk (R', R_z) and Legitimate Measures (L): Total Group 29

4.1 Split-half Reliabilities of Vocabulary Test (Sample Sizes in Parentheses) 36

4.2 Mean RTOOE Score (Standard Deviation in Parentheses) 36

4.3 Mean Vocabulary Test Score for Conventional Groups 37

4.4 Mean Vocabulary Test Score for Coombs Groups 37

CHAPTER I: INTRODUCTION

PROBLEM

Risk taking behavior has become an increasingly important variable of concern in educational and psychological research. It enters into such basic research areas as motivation, personality, decision making, etc. Indeed, two leading researchers in risk taking have recently published a book entirely devoted to matters such as these (Kogan and Wallach, 1964).

On the more applied side, it appears that risk taking may play a part in vocational choice (Ziller, 1957a) and in curriculum choice in college. In addition, risk taking behavior is an important characteristic of certain professions; e.g., scientists must conjecture, military and business men must take "calculated" risks, etc. Therefore, in vocations such as these, measures of risk taking might be useful aids in selection and/or classification of personnel, and in the evaluation of curricula. Several studies (e.g., Torrance and Ziller, 1957; Williams, 1961) have concerned themselves with this area.

Finally there is evidence (Hammerton, 1965; Quereshi, 1960; Sherriffs and Boomer, 1954; Slakter, submitted; Votaw, 1936) that for objective examinations with penalties for incorrect responses, risk taking behavior affects test score. Therefore, it may prove possible to increase the validity of these objective examinations for certain purposes, by adjusting the test score for risk taking tendencies.

Interest in this project centered on the latter type of risk taking; i.e., risk taking on objective examinations (RTOOE). This behavior as defined here, is limited to objective tests which indicate in their directions that there is a penalty for incorrect responses. Under the above conditions, RTOOE is the tendency to respond to an item when the correct response is not known with certainty.

The measurement of RTOOE provides an objective, disguised measure of risk taking behavior, and as such appears to be a potentially valuable measure in the research areas mentioned previously. In brief, this project was concerned with three aspects of RTOOE: (1) its measurement, (2) its generality, and (3) its effect on test score.

SELECTIVE REVIEW OF LITERATURE

Measurement of RTOOE

The following is a brief description of the measures of RTOOE, and their limitations (Slakter, 1967, pp 33-36)*.

> Wiley and Trimble (1936) appear to be among the first to have demonstrated the feasibility of using objective tests to measure personality traits. They administered 4 objective tests to 59 students in a general psychology course as part of the regular achievement testing program. *S*s were instructed to answer the questions and to make each answer with a +, ?, or √ to represent respectively "certainty," "doubt," and "guess," with respect to the degree of sureness involved in responding to each question. In addition, *S*s were advised "you will be penalized for inconsistencies; i.e., if your answer in any given case is wrong while your mark indicates you are positive it is right, etc." Each test was scored for the number of +'s, ?'s, √'s, and achievement. The authors found that the scores for "certainty," "doubt," and "guess" were more reliable (equivalence and stability) than the achievement scores. However, no attempt was made to label the trait or traits being measured.
>
> Swineford (1938), in a follow-up of the Wiley and Trimble (1936) study, administered a true-false test to 160 college students, and directed them to mark their answers 4, 3, or 2 to signify the degree of sureness (where "4" indicated certainty). *S*s were also instructed that there would be a penalty of double the amount claimed for incorrect answers. Swineford referred to the trait measured as the tendency to gamble. The

---

*Reprinted from the *American Educational Research Journal* by permission of the American Educational Research Association. Copyrighted 1966.

reliabilities (split-half) for the "4's" and "3's" were again higher than that for achievement, with the highest reliability for "4's." Therefore, Swineford defined her index of RTOOE, R_s, as

$$R_s = \frac{\text{number of errors marked "4"}}{\text{total errors} + u/2} ,$$

where u is the number of items unattempted by the examinee. In Swineford's study, R_s was found to be uncorrelated with achievement test score.

In a later study (1941), Swineford administered 4 different objective tests (non-language figure test, general information, vocabulary, logical deduction) to 457 high school students with her special directions. Approximately 25 percent of the *S*s were eliminated from the study "either because on one or more of the tests no extra credits were requested, or because on one or more of the tests no errors were made among the items attempted." For the remaining *S*s it was found that males had higher RTOOE scores than females, the distribution of R_s was positively skewed, and R_s was uncorrelated with aptitude test score. Interestingly enough, this writer could find no further studies of RTOOE using R_s. Possible reasons for this may include the awkwardness of directions, and other limitations to be discussed shortly.

Another measure of RTOOE that has been used with conventional directions where the examinee is apprised of a penalty for incorrect responses is the number of unattempted items. In order to have high scores indicate high risk taking, the measure will be defined here as

$$R_u = (N - u)/N ,$$

where N is the number of items on the test. Studies using this measure of RTOOE include Votaw (1936), and Sherriffs and Boomer (1954). However, both of these studies were actually more concerned with the problem of RTOOE affecting test score rather than with measuring RTOOE as such.

Still another measure of RTOOE for use with conventional directions with a penalty for incorrect responses was described by Ziller (1957a), and used in studies by Ziller (1957b), Torrance and Ziller (1957), and Stone (1962). Ziller's index, R_z, is defined as

$$R_z = \frac{\frac{n}{n-1} w}{\frac{n}{n-1} w + u} ,$$

where n is the number of alternatives, and w is the number of incorrect responses.

Coombs (1953, 1956) has suggested a variant of the conventional directions which incorporates a penalty for incorrect answers. Coombs' directions essentially instruct the *S*s to cross out all the alternatives which they consider wrong (as opposed to the conventional crossing out of the correct alternative). Therefore with the Coombs' directions, attention is focused not on the RAS, but on the complement of the RAS.* For a n-alternative item, the *S* receives 1 point for each incorrect alternative in RAS' (the complement of RAS), and minus (n-1) points for each correct alternative in RAS'. Danielson (1956) utilized Coombs' directions to investigate RTOOE, and defined his index, R_d, as

$$R_d = \frac{(n - 1)\text{ (number of correct alternatives in the RAS')}}{\text{number of incorrect alternatives not in the RAS'}}$$

In examining R_z, R_d, and R_s it becomes apparent that all three suffer from the following limitations:

---

*The complement of the RAS is the set containing all alternatives that the examinee has eliminated from consideration as the correct response.

1. The denominators for some *S*s may equal zero, resulting in an undefinable score for the *S*s and their elimination from the study. In general these *S*s may tend to be high achievers.

2. The denominators have different values for different *S*s, resulting in different reliabilities.

3. There is no way to determine whether an item is correct by guessing (i.e. RAS: A, B, C, D, E, F, G)*; or whether the correct response indicates complete information (RAS:H)*.

4. There is no way to determine whether an item is incorrect by guessing (i.e. RAS: A, B, C, D, E, F, G, L, M, N, O)*; or whether the incorrect response indicates misinformation without guessing (RAS: I, J, K)*.

With respect to R_u, it is apparent that information is an important confounding factor here. Quite conceivably a low risk taker with much information could have fewer omissions (few omissions imply high RTOOE) than a high risk taker with little information.

With the above deficiencies in mind, the author attempted to devise a new measure that would reduce or eliminate these limitations. The new index relies on the Coombs' type items used by Danielson (1956), and is based upon a comparatively few "nonsense" questions imbedded in "legitimate" achievement or aptitude questions. A "legitimate" question is defined as a question that has one correct (or best) alternative, and (n-1) incorrect alternatives for the given population. A "nonsense" question is one that has no correct (or best) alternative, and no incorrect alternatives for the given population. For example, a nonsense question suitable perhaps for use in an introductory course in educational psychology might be:

---

*These symbols are not necessary to an understanding of the text. However, for an explanation of these symbols, see Slakter (1967, p 32).

What is the child's age in months when the Lumdane Effect is usually first noticed?

a. 1-2
b. 3-4
c. 5-6
d. 7-8

For the given population (and indeed for all populations to the best of the author's knowledge) this question has no correct alternative and no incorrect alternatives, and is therefore, a nonsense question.

The index R, based upon the nonsense questions, is defined as

$$R = \frac{\text{the number of alternatives crossed out in the nonsense items}}{q(n - 1)},$$

where n is the number of alternatives, and q is the number of nonsense questions. Limitations 1 and 2 above are now eliminated. In addition, it would appear that this new measure is more purely risk taking and less contaminated by misinformation, information, and chance. In other words, limitations 3 and 4 appear to be reduced.

Swineford and Miller (1953) report data that is relevant to the new measure of RTOOE. In their study Swineford and Miller compared the effects of directions regarding guessing upon the characteristics of the test and the items, and upon guessing itself. They used a 100-item vocabulary test which included 20 nonsense questions as defined in this paper. Their major emphasis, however, was on guessing and not risk taking (only one form of direction included a penalty for incorrect responses), and their major interest appeared to be with the effects

> of the different directions on item or test statistics; i.e., the biserial correlations between the items and total test score. However, they do provide data from the group whose directions indicated a penalty for incorrect responses, which enables a partial investigation of R', where we define
>
> $$R' = \frac{\text{the number of nonsense items attempted}}{\text{the total number of nonsense items}}$$
>
> It is seen that R' as defined here is R for conventional directions with a penalty for incorrect responses.
>
> Swineford and Miller were not primarily interested in RTOOE as such, and hence did not calculate the reliability of R'. They did report, however, what might be interpreted as a split-half reliability of .954 between two groups of 10 nonsense items each. Therefore, with the Spearman-Brown formula, the estimate of the reliability of the 20 items is .976. In addition, the mean and standard deviation for the number of responses to the nonsense questions is given as 11 and 8.14, respectively. These latter two statistics can be used in calculating a reliability coefficient by the Kuder-Richardson formula (21), which turns out to be .974. For the sake of comparison, the 80 item vocabulary test (composed of legitimate items) had a reliability coefficient of .856 by the Kuder-Richardson formula (21).

Several months after the publication of the above material, another extremely important reference was discovered while searching the literature on a related topic. The paper (Granich, 1931) not only appears to be the earliest to describe a measure of RTOOE, but of even more significance, is that the measure of risk "...is obtained by scattering, among the genuine questions of a true-false or multiple choice examination, a number of questions which involve very obscure facts, or newly coined names." (Granich, 1931, p 145). Here then, is the first use of nonsense questions to measure RTOOE. Granich found no relation between achievement and RTOOE, and suggested many interesting possibilities for further investigation: e.g., "a study of the distribution of student-types as regards guessing." Unfortunately, there is no evidence of other work by Granich along these lines, and indeed, this important study appears to have remained almost completely unnoticed in the literature related to RTOOE.

Finally, in a study examining R_z, R_u, R_d, and R (Slakter, 1967), it was found that the measures of RTOOE were reliable in the split-half sense, with R achieving reliabilities greater than .80 with just 6 items. In addition, the correlation matrix of risk, aptitude, and achievement measures offered some evidence for convergent and discriminant validity.

Generality of RTOOE

Only one study was found in the literature dealing with the generality of RTOOE as defined in this study. There were studies that dealt with error scores on speeded tests (Fruchter, 1950; Thurstone, 1938), but this behavior appears to the writer to be something other than risk taking, and does not fall under the definition of RTOOE as used here. There have been several studies concerned with the generality of risk taking itself (e.g., Slovic, 1962), but again these are not relevant to the subject of interest.

The single study relating to the generality of RTOOE was one referred to previously (Swineford, 1941), in which the following tests were used: non-language geometrical figures, general information, vocabulary, and logical deduction. Correlations among the RTOOE scores for the tests ranged from .2 to .8, with the highest correlation occurring between the RTOOE scores on the verbal tests (general information, vocabulary). The author states (Swineford, 1941, p 442): "Examination of the correlations reveals that a G factor common to all the tests and an overlapping factor between the verbal tests may be postulated."

Effect of RTOOE on Test Score

Quereshi (1960) has shown that factors such as risk taking play an important part in mental test scores. One might therefore question whether there is some optimum risk taking strategy that will maximize average test score for examinations that include a penalty for incorrect responses. Varty (1946, p 211) with respect to this problem, declared: "It is the writer's opinion that over a relatively long period of time or a large number of items, the individual student will profit by answering all items which are not in the 'pure guess' classification." This seems like intuitively sound advice, but there still remains the problem of determining which items fall into the "pure guess" category; e.g., is it possible that two people with the same information on an item will classify the item differently with respect to "pure guess"? If so, perhaps a better strategy is to answer all questions, "pure guess" or not.

With respect to conventional directions, various studies (Hammerton, 1965; Sherriffs and Boomer, 1954; Slakter, submitted; Votaw, 1936) have found that at least some types of students would score higher if they responded to all items. With Coombs type directions, however, the writer is unaware of any evidence concerning optimum strategy.

OBJECTIVES

Measurement of RTOOE

Preliminary studies provided evidence that R is a promising measure of RTOOE, but unfortunately the Coombs type directions used are unfamiliar to most test takers. Therefore, it is reasonable to be interested in R', where R' is a measure of RTOOE which uses nonsense questions under regular "do-not-guess" directions (see page 8). The specific area of interest originally was with the various correlations among R_d, R_u, R_z, R, and R'. However, this phase of the project was modified in view of two older studies found in the literature, which the writer was unaware of at the time the proposal was submitted. These two studies have already been referred to (Swineford and Miller, 1953; Granich, 1931), and their implications and the revised problem will be discussed in Chapter II.

Generality of RTOOE

If an individual tends to take risks on an achievement test in mathematics, will he also tend to take risks on an achievement test in language? In other words, is RTOOE highly specific to the content or the type of test, or does it tend to be a more general trait. In particular, for this study, consider mathematical and language aptitude tests, and achievement tests in mathematics and language. What are the correlations among the RTOOE scores on these tests?

Effect of RTOOE on Test Score

In order to study the effect, if any, of RTOOE on average test score, an experimental study was designed. In the experiment, RTOOE was manipulated, and the resultant effect on average test score noted. One facet of this phase was concerned with conventional do-not-guess directions, and another facet utilized Coombs type directions. Specifically:

1. For conventional do-not-guess directions, which of the following strategies results in highest average test score?
   a. Answer all items that you "think" will help you to obtain as high a score as possible. (usual risk)
   b. Answer all items. (high risk)
2. With respect to Coombs type directions, which of the following strategies result in the highest average test score? (Assume 4 alternatives for each item.)
   a. Always select exactly one alternative. (low risk)
   b. Always select exactly two alternatives. (medium risk)
   c. Always select exactly three alternatives. (high risk)
   d. Always select those alternatives, and only those alternatives that you are sure are wrong. (usual risk)

SECURING OF SAMPLE

As the first step in obtaining *S*s for the study, the Director of the Western New York School Study Council was consulted for the names of school systems in the vicinity that might be interested in participating in the project. The names of 6 suburban systems were suggested, their superintendents telephoned by the principal investigator, and the nature of the project briefly explained to them. Follow up letters were then sent out to the superintendents in order to provide more information upon which to base a decision. Five of the six superintendents indicated interest in participating, but one system of the five was ultimately eliminated because of scheduling difficulties.

In addition, the superintendent of a large Western New York city school system was approached by telephone. The superintendent indicated that an assistant superintendent would consider the matter, and would make a decision shortly after receiving the follow up letter. Several months passed without word, and therefore the assistant superintendent was telephoned and reminded of the project. A decision was again promised shortly, but another telephone conversation promising a decision shortly, and a passage of several more weeks, made it apparent that the decision had been "no decision." Therefore, it was regretfully decided not to press any further for inclusion of the city system into the project.

To counterbalance the comparisons lost by the unavailability of the Western New York city system, an opportunity arose to include a Canadian city system. A principal in the Canadian system, a doctoral student in a class of the principal investigator, was approached, and through him an initial contact was made with the superintendent of public schools. The superintendent proved to be interested in the project, and two Canadian schools, one English speaking and one French speaking (or bilingual), were included in the study.

The final sample, therefore, was supplied by four suburban school systems in Western New York State, and one Canadian school system. All of the New York systems participated in both the generality and effects of RTOOE studies, while the Canadian system took part only in the generality phase.

The 8th grade was selected for the generality study in order to have a group that was young enough so that the drop-out problem would not appreciably bias the results, and yet old enough so that they would be able to understand the directions and withstand approximately two hours of testing. The 11th grade was chosen for the experimental study on the effect of RTOOE on test score, since the directions for some of the treatments were quite unconventional, and the 12th grade students were administratively unavailable.

In three of the suburban, and one of the Canadian schools, the 8th grade students tested were selected at random from the entire class. In each case, a random sample of 100 was drawn; the entire class numbered 459 in the largest, 176 in the smallest. In one school, it was decided to test the entire 8th grade class, since there were only 145 students. Finally, in one of the Canadian schools whose 8th grade was split administratively into two sections, one entire section of 128 was selected at random.

In the experimental study at the 11th grade level, *S*s were always assigned at random to each of the six treatments. In two of the schools, the entire class (377 *S*s; 131 *S*s) took part in the study; in one of the schools, an intact group of 159 participated. In the largest 11th grade class, a random sample of 150 was selected from the total group of 468. It is important to note that in both the 8th grade testing, and the 11th grade testing, there was natural attrition due to absentees, etc. Therefore, the actual number tested was in general slightly lower than the figures listed above.

COMMUNITY CHARACTERISTICS

The four New York State school systems, while all within a 45 minute automobile ride of the large city, reside in somewhat different communities. One system serves a small (6,000 population) middle and upper-middle class residential village. Of the group, this community has the highest median education and income level. Another system is supported by a more populated (84,000) community, somewhat closer to the city, and somewhat industrialized. A third system is found in a large village (26,000), which is mostly residential or rural, with some industry. The last community may be characterized as a small rural village (3,000).

The Canadian city has a population of approximately 40,000, with the basic industry being steel and steel products. The inhabitants are mostly second and third generation immigrants, with a comparatively large French-Canadian population, residing chiefly in a separate sector of the city.

TEST CONSTRUCTION

The testing in the public schools required the construction of five tests, four in the generality study, and one for the study of the effect of RTOOE on average test score. The four tests in the generality study had to include nonsense questions as well as legitimate questions; the single test for the effect of RTOOE on average test score consisted only of legitimate questions.

The construction of the achievement tests for the generality study originated with an examination of the New York State syllabus (State University) and its Canadian counterpart (Ontario Department). An attempt was made to have the legitimate items that would be common to both syllabii at the 7th grade level. The 7th grade level was selected because the testing was to be completed early in the school year, and it was thought that 7th grade material would be appropriate for the achievement tests. Included with the legitimate questions for each achievement test, were 10 nonsense questions. One test was constructed to "measure" language, and another arithmetic achievement. (see appendix C).

In the construction of the aptitude tests, an attempt was made to construct items that the *S*s would be less familiar with, but perhaps would be able to solve correctly. In many respects, however, the achievement and aptitude tests in a similar dimension (e.g., mathematical) were of quite similar nature. Indeed, it is often quite difficult to determine where an achievement test ends and where an aptitude test begins. It was hoped, however, in each case, that the identifying title of the test might help the examinee to make this determination. As in the achievement tests, each aptitude test included 10 nonsense questions. One test was constructed for language aptitude and another for mathematical aptitude. (see appendix B).

Next the tests were tried out on two ninth grade classes in a local summer school to determine if the directions were clear, and the items, both nonsense and legitimate, suitable. Ninth grade classes were used because they were convenient, and since the vast majority of the students were in summer school for remedial work, the level was deemed acceptable. The mathematics aptitude and achievement tests were administered to a 9th grade mathematics class; the language aptitude and achievement tests were administered to a 9th grade English class. In each case, following the results, changes were made in the directions and items.

It was decided that the 11th grade test would be a vocabulary test for ease of construction, and hence word lists appropriate for the 11th grade level were examined (Thorndike & Lorge, 1944). An item pool of approximately 150 items was constructed, and this list reduced to 100 items of varying (a priori) difficulty. In addition, much time was spent attempting to insure that the directions for the six different treatment groups were as clear and simple as possible.

The test was then administered to a small group of graduate students and secretaries, and slight revisions of items and directions followed. Copies of the 100-item test, and the set of six different directions are found in the appendix (E, F).

ADMINISTRATION OF TESTS

The testing in the public schools was accomplished during the months of October and November in 1966. The 8th grade classes were tested in one large room in the school, generally the cafeteria. Directions for the administration of the tests are found in the appendix (A).

The 11th graders were split into the following three groups on the basis of similarities in their directions:

1. Conventional directions

    a. usual
    b. answer all

2. Coombs

    a. select one alternative
    b. select two alternatives
    c. select three alternatives

3. Coombs

    a. usual

Each group was then assigned to an individual room. This procedure was followed in every school except one. In the latter school, three distinct rooms were unavailable because of space problems, and therefore the testing was carried out in different sections of the auditorium. The directions for the administration of the test are given in the appendix (D).

In general, the cooperation from the schools ranged from good to excellent, and the students appeared to complete the testing sessions, which for the 8th graders was quite lengthy, in fairly good spirits. Subjects were notified that the scores were for research purposes, and would not in any way affect their grades, acceptance by college, etc. They were asked, however, to try to do their best, and the impression gained by the examiners was that the students were generally quite sincere in their efforts. Clearly though, the results of this project are limited by these conditions under which the tests were taken. However, it was felt by the principal investigator, that it might be somewhat unethical to deceive the subjects as to the possible uses of their test scores. Furthermore, one might strongly suspect that the school administration might have been less than enthusiastic if the students were to be led to believe that the scores would enter into their grades, etc.

REMAINING CHAPTERS

In the chapters that follow, the three facets of the study will be examined in more detail. Chapter II will describe the original and revised problem in the measurement of RTOOE. Chapter III is concerned with the generality of RTOOE, while Chapter IV deals with the experimental study on the effects of RTOOE on test score. In the fifth and final chapter, the entire project and its findings will be summarized, and certain conclusions advanced.

CHAPTER II: MEASUREMENT OF RTOOE

ORIGINAL PROBLEM

Judging from the results of early studies (Slakter, 1967), the nonsense items appeared to provide a promising basis for the measurement of RTOOE. However, the use of the nonsense items with the unfamiliar Coombs type directions was definitely limiting. Therefore, it was of interest to investigate the characteristics of the nonsense questions when used with conventional directions; i.e.,

$$R' = \frac{\text{the number of nonsense questions attempted}}{\text{the total number of nonsense questions}} .$$

The original problem was concerned with the values of the various correlations among R_d, R_u, R_z, R, and R'. However, when the writer discovered two studies in the literature that he had been unaware of at the time of the original proposal, he decided that the evidence for the use of R' was already somewhat established. Therefore, this phase of the study became of less interest, and the corresponding problem was revised extensively. In the next section, the implications of the two additional studies will be examined, and a following section will state the revised problem.

ADDITIONAL STUDIES RELEVANT TO R'

Granich (1931) appears to have developed the first measure of RTOOE, although in the spirit of that period, he referred to it as an "index of guessing." The index was obtained (Granich, 1931, p 145) "...by scattering, among the genuine questions of a true-false or multiple choice examination, a number of questions which involve very obscure facts, or newly coined names." At the time, Granich was more interested in reducing guessing than in measuring RTOOE. In his study, Granich examined the effects of two types of directions on guessing. One type of direction was the conventional do-not-guess; the other was the conventional do-not-guess with an added warning that the examination contained nonsense questions. However, Granich certainly recognized the potential of the measure of RTOOE when he wrote (Granich, 1931, p 152):

> The greatest utility of the technique lies in further theoretical investigation. Among the possibilities are: (1) Correlative work and the determination of relations between guessing and the different elements of a test situation. (2) Comparison of the amount of guessing going into different types of tests, including even completion tests: also, estimation of the relative amount of guessing going into specific tests. (3) A study of the distribution of student-types as regards guessing. (4) Analysis of characteristics of questions which invite guessing. (5) Study of the extent to which students are aware of their guessing. (6) Retention of content with different tests. (7) A measure of one phase of classroom discipline, and of the effectiveness of different instructions, motivations, personalities, etc. (8) Experimentation leading to the perfection of the index, confirming its validity, and determining its possible application to the classroom. The possibility of employing it as a refinement in scoring is of greatest importance here.

Although Granich did not comment upon the reliability of his index, a frequency distribution of the number of responses to the nonsense items was supplied in a table (Granich, 1931, p 152). From the table, the mean and standard deviation for the number of responses by the conventional do-not-guess group was calculated to be 4.21 and 2.506 respectively. These values were then substituted in the Kuder-Richardson formula #21, and the reliability was estimated to be .68. This reliability, while quite high for a 10-item test, is lower than comparable values found for R.

However, some 22 years after the paper by Granich, Swineford and Miller (1953) reported another study utilizing nonsense questions. The main interests of the Swineford and Miller investigation appeared to be with the effects of guessing on test score, test statistics, and item statistics. They were also interested in the effects of directions upon guessing, and the relation between guessing and performance. Of their three experimental groups, however, only one received do-not-guess directions, and therefore this is the only group of interest in the study of RTOOE. The examination used in the experiment was a vocabulary test, constructed to include 80 legitimate questions, 10 nonsense questions, and 10 "extremely difficult" (Swineford and Miller, 1953, p 129) questions. By our definition, the "extremely difficult" questions were actually nonsense questions, since the authors state (Swineford and Miller, 1953, p 129): "...it is unlikely that any of these words would be familiar to the average college graduate."

The data resulting from the Swineford and Miller study reinforce the contention that the "extremely difficult" items were actually nonsense items. For the group receiving the do-not-guess directions, the mean number of responses to the nonsense questions was found to be 5.48 versus 5.52 for the "extremely difficult" items; the standard deviation for the number of responses to the nonsense items was calculated to be 4.13 versus 4.10 for the "extremely difficult" questions. Furthermore, the correlation between the two sets of 10 items was found to be .954, which led the authors to point out (Swineford and Miller, 1953, p 134): "Correlations of this magnitude, based on so small a number of items, indicate an unusually high degree of consistency of response. Reaction to the difficult items was essentially the same as reaction to the nonsense items." Finally, using the mean and standard deviation for the set of twenty items, nonsense and "extremely difficult," a measure of reliability by the Kuder-Richardson formula #21 was calculated to be .974. With the above evidence, therefore, it seems reasonable to assume that the "extremely difficult" items were actually nonsense items for the group tested. Note that this conclusion justifies Granich's use of "very obscure facts, or newly coined names" (Granich, 1931, p 145) for the basis of his nonsense questions.

Examining the results of the Granich study (1931), and the Swineford and Miller study (1953), we have seen evidence demonstrating the reliability of R'. Unfortunately, neither study provided any information on the validity of R'. However, some evidence on the validity aspect was found as part of a larger study (Slakter and Cramer, in preparation), completed independently of this U. S. Office project. Therefore, let us consider the revised problem described in the following section.

REVISED PROBLEM

Of all the measures of RTOOE based on legitimate items, the one which has been studied most thoroughly is R_z. Therefore, if we consider R_z as a standard, it is of interest to find how R' is related to R_z. More specifically, what is the correlation between R_z and R'?

RELATION BETWEEN R_z and R'

As part of a larger study concerned with the relation between vocational choice and risk taking, entering freshmen in the fall semester of 1966 at the State University of New York at Buffalo (SUNY/B) were tested during the summer of 1966. During a 15 minute segment of the freshmen testing period, the COLLEGE ENTRANCE VOCABULARY TEST (CEVT) was administered to all *S*s. The CEVT is supposedly a legitimate synonym-antonym type vocabulary test, but is actually composed of 20 nonsense items embedded in 40 legitimate items. The directions for the CEVT indicate that there is a penalty for incorrect responses, so that the 20 nonsense items furnished a basis for the calculation of R', while the 40 legitimate items provided a basis for the calculation of R_z. In other words, each *S* received two RTOOE scores, one by R', and the other by R_z. The correlation between these two measures of RTOOE, therefore, provided an estimate of concurrent validity. Complete data were obtained on 1,010 males and 776 females, for a total of 1,786 freshmen.

Table 2.1 presents the means and standard deviations of R', R_z, and V (the vocabulary score on the 40 legitimate items, "corrected" for guessing) for the males, females, and total group. It is seen that the females scores slightly higher than the males on R_z, while the males scored slightly higher than the females on

Table 2.1

Means and Standard Deviations for
Vocabulary and Risk Taking Measures

	Males	Females	Total
N	1,010	776	1,786
$\bar{X}_{R'}$	.56	.55	.56
$\bar{X}_{R_z}$	.72	.74	.73
$\bar{X}_V$	24.4	27.6	25.8
$s_{R'}$	.37	.36	.37
s_{R_z}	.28	.29	.29
s_V	7.03	6.46	6.97

R'. However, the strength of relation* between sex and R' was estimated to be .001; the strength of relation between sex and R_z was estimated to be .000. Furthermore, the mean differences for sex on R' and R_z were each nonsignificant at the .05 level. It is clear, therefore, that the relation between sex and RTOOE, as measured by R' or R_z on the CEVT, was negligible.

In addition, the values found for the mean of R' for this freshmen group at SUNY/B appear to be quite comparable to the mean values found for R with undergraduates and graduate education students at the University of California at Berkeley (Slakter, 1967, p 39). For males, the mean of the freshmen group (.56) was slightly higher than the undergraduate mean (.51), and the graduate mean (.50). However, for females, the freshmen mean (.55) was slightly lower than the undergraduate mean (.56), and the graduate mean (.61). For the total group, the freshmen mean was .56, as compared to the undergraduate mean of .55, and the graduate mean of .53. Considering that R and R' are different measures, based on different items, given to different types of students, these comparable results are impressive.

The standard deviation of R' for the freshmen group at SUNY/B appears to be slightly larger in each case, than the standard deviation of R for the undergraduate and graduate Berkeley students; i.e., for males .37 versus .22 and .32 respectively, for females .36 versus .22 and .25, and for the total group .37 versus .22 and .32. The tendency for the Berkeley groups to have smaller standard deviations might be explained on the basis that the Berkeley groups were more homogeneous in RTOOE, since they tended to be in one curriculum (education).

In similar fashion, values found for the mean R_z on the CEVT for the SUNY/B freshmen, are quite comparable to the corresponding values for the Berkeley groups on part 1 of the Concept Mastery Test; i.e., for males .72 versus .72 and .80 respectively, for females .74 versus .62 and .78, for the total group .73 versus .64 and .80. Once again, the standard deviations were slightly larger in the freshmen group; i.e., .28 for males versus .24 and .25 respectively, .29 for females versus .23 and .28, and .29 for the total group versus .23 and .25.

Not too surprisingly, the females scored higher than the males on the legitimate vocabulary portion of the CEVT. The estimate of the strength of relation between sex and vocabulary was .052, and the mean difference was significant at the .05 level.

---

*For a description of strength of relation, see Hays (1963, pp 323-32).

Table 2.2 provides the split-half reliabilities ("corrected" by the Spearman-Brown formula) for the vocabulary and RTOOE measures. It is seen that there were very slight differences in reliability between the males and females, and that R' appears to have been more reliable than either R_z, or the legitimate vocabulary score. The reliability of R' for the SUNY/B freshmen (.96) compares favorably to the reliability of R for the Berkeley groups (.95 for each group when the Spearman-Brown formula was used to estimate the reliability of a 20-item test). The reliability estimate of .96 for R' also is about the same magnitude as that calculated for R' (.98) from the Swineford and Miller data (1953), and both values are higher than that (.88) found from the Granich data (1931). The latter value, like the other reliability values, required the use of the Spearman-Brown formula to estimate the reliability of a 20-item test.

Table 2.2

Split-half Reliabilities* for Vocabulary and Risk Taking Measures

	Male	Female	Total
R'	.96	.96	.96
R_z	.72	.66	.70
V	.69	.68	.70

*"corrected" by Spearman-Brown formula

The correlation matrix for the vocabulary and RTOOE measures is presented in Table 2.3. A fairly high correlation between R' and R_z was found for males, females, and the total group. When the reliability of R_z is taken into consideration, the magnitude of the correlations between R' and R_z appears even more impressive. For example, from the reliability of R_z for the males, we find that approximately 72% of the variability in R_z was non-error; from the square of the correlation between R' and R_z for the males, we see that approximately 64% of the variability in R_z could be predicted from R'. Therefore, it is clear that practically all of the reliable variance in R_z for the males, was predictable from R'. The same is essentially true for the females, although to a somewhat lesser extent.

Table 2.3

Correlations for the Vocabulary and Risk Measures

	Males		Females		Total	
	R_z	V	R_z	V	R_z	V
R'	.80	-.05	.71	-.07	.76	-.06
R_z		.01		-.10		-.03

It is also important to note that both R' and R_z were essentially uncorrelated with vocabulary for the males, females, and total group. For these *S*s, it is clear that RTOOE, as measured by the CEVT, had little linear relation with vocabulary.

In conclusion, the studies by Granich (1931), and Swineford and Miller (1953), provided data attesting to the reliability of R'. The results of the CEVT study on the SUNY/B freshmen demonstrate, that in addition to providing a reliable measure, R' appeared to measure essentially the same dimension as that of R_z. Therefore, it is reasonable to assume that R', as well as R, can be used to measure RTOOE. Finally, it will be seen in Chapter III, that there is even stronger evidence for the use of R' as a measure of RTOOE.

CHAPTER III: GENERALITY OF RTOOE

INTRODUCTION

Results of various studies (e.g., Stone, 1962; Swineford, 1941; Slakter, 1967) indicate that the RTOOE behaviors of individuals on a given test are impressively consistent. Generally speaking, no matter which index of RTOOE is used, the measure of RTOOE provides reliabilities at least as high as those provided by the aptitude or achievement dimension that the test was designed to assess. As a next step, therefore, one wonders whether RTOOE behavior is consistent across various kinds of test situations; e.g., different types of achievement and aptitude tests, different types of items, varying difficulties, etc. In other words, are the RTOOE behaviors of individuals general across various kinds of testing situations, or are these RTOOE tendencies highly specific to the particular testing situation?

For this study, the question of concern was the generality of RTOOE across different types of tests. Specifically, interest was centered on the values of the RTOOE correlations among the following four types of tests: (1) language aptitude, (2) mathematics aptitude, (3) language achievement, and (4) mathematics achievement.

In the only previous study of the generality of RTOOE that the writer is aware of, Swineford (1941) administered four different tests to 457 high school freshmen. The four tests were described by Swineford as follows (1941, p 439):

> One is a non-language test, Paper Form Board, in which each of the twenty-eight items consists of one of four geometrical figures cut into three or four sections. The subject is to determine which figure the sections would fit if they were reassembled. The second test, General Information, is a multiple-choice test of one hundred items covering factual information. The third is a fifty-item multiple-choice test of vocabulary. The fourth is a thirty-six item true-false test of logical deduction based on series of inequalities written in terms of letters of the alphabet.

Unfortunately, however, because of the nature of the RTOOE index or the directions, many of the *S*s had to be eliminated from the study. Swineford (1941, p 439) stated: "Of the four hundred fifty-seven pupils who were tested, seventy-four boys and thirty-nine girls were eliminated from the gambling study either because on one or more of these tests no extra credits were requested, or because on one or more tests no errors were made among the items attempted."

Swineford (1941) found that males displayed higher RTOOE than females, the distribution of R_s was relatively uncorrelated with legitimate test score. With respect to RTOOE correlations among the four tests, Swineford (1941) found that the correlations for male, female, and total group ranged from about .2 to .8, with little, if any, sex difference. Swineford concluded (1941, p 442): "Examination of the correlations reveals that a G factor common to all the tests and an overlapping factor between the verbal tests may be postulated."

*S*s

*S*s were students in the 8th grade, selected from four schools in western New York State, and two schools from a city in Ontario, Canada. In one of the Canadian schools, classes were conducted only in English; in the other Canadian school, both French and English were spoken. In three of the New York State schools, and the bilingual Canadian school, a random sample of 100 *S*s was selected from the entire 8th grade class. In the remaining New York State school, it was decided to test the entire 8th grade class of 145 students. In the English-speaking Canadian school, since the 8th grade was split administratively into two large sections, one entire section of 128 students was selected at random. As expected, in each case there was attrition due to absences, etc., so that the sample size in any particular school was smaller than that listed above. The exact sample sizes are found in the results section.

PROCEDURE

*S*s were seated in one large room, and the first of two test batteries administered to them was the STANDARD EDUCATIONAL INTELLIGENCE TEST (SEIT). The general directions for SEIT began with the sentence: THIS BOOKLET CONTAINS A TEST WHICH WILL GIVE YOU A CHANCE TO SHOW WHAT YOU KNOW AND HOW WELL YOU THINK. SEIT was composed of two parts, the first supposedly a measure of language aptitude, and the second supposedly a measure of mathematical aptitude. The language section contained 10 nonsense items embedded in 40 legitimate items; the mathematics section was composed of 10 nonsense items embedded in 30 legitimate items. In constructing the legitimate items for both parts of SEIT, an attempt was made to select items that would be less familiar to the *S*s than those legitimate items on the corresponding achievement tests. One measure of the success of this attempt can be determined by a comparison of the difficulties of the aptitude and achievement tests in either the language or mathematics dimension. The comparisons of these difficulties are examined in the discussion section.

The second test battery administered to the *S*s was titled STANDARD EDUCATIONAL ACHIEVEMENT TEST (SEAT), and the general directions began with the sentence: THIS BOOKLET CONTAINS A TEST OF SOME OF THE KNOWLEDGE YOU HAVE GAINED DURING YOUR SCHOOL YEARS. Like SEIT, SEAT was composed of two parts, the first supposedly a measure of language achievement, and the second supposedly a measure of mathematical achievement. The language section in the SEAT booklet contained 10 nonsense questions embedded in 30 legitimate questions; the mathematics section of SEAT also consisted of 10 items embedded in 30 legitimate items. In both the language and mathematics sections of SEAT, an attempt was made to have all of the legitimate items cover material that the *S*s had previously been exposed to in their formal schooling.

Therefore, the administration of SEIT and SEAT provided RTOOE and legitimate scores on the following four types of tests: (1) language aptitude, (2) mathematics aptitude, (3) language achievement, and (4) mathematics achievement. Each of the four tests provided a measure of RTOOE on nonsense items (R'), a measure of RTOOE on legitimate items (R_z), and a legitimate score (L).

The administration time for SEIT and SEAT was 45 minutes each, for a total of 90 minutes. With an additional 10 minutes for passing out the tests, etc., the total testing time was approximately 1 hour and 40 minutes. In all cases except one, the vast majority of *S*s had ample time to finish the tests. In one instance, the testing time had to be shortened by approximately five minutes, and the differences, if any, will be noted in the discussion section. Copies of SEIT and SEAT are included in the appendix (B, C).

RESULTS

Table 3.1 presents the means and sample sizes for the six schools on the risk and legitimate measures. The legitimate scores (L) were calculated as the number right minus one-third the number wrong. Males, females, and totals are displayed separately. The English-speaking Canadian school is labeled S_4; the bilingual Canadian school is labeled S_5. School S_6 designates the New York State school that was stopped approximately five minutes early, and thus had less time to finish than the other schools.

The distributions of R' and R_z were examined for each of the four tests in each of the six schools. In each case, the distribution of the RTOOE measure was negatively skewed, with the mode at the extreme right.

The correlations among the risk and legitimate measures for the six schools are provided in Table 3.2. The diagonal entries in the table are the split-half reliabilities (odd versus even, "corrected" by the Spearman-Brown formula). From a visual comparison of the correlation matrices for the males and females, it was apparent that the sex differences were negligible. Therefore, Table 3.2 contains only the results for the total group.

DISCUSSION

The finding that sex differences in the correlation matrices were negligible confirmed similar results found by Swineford (1941). An examination of Table 3.2 revealed impressive evidence, across all six schools, for convergent and discriminant validity as described by Campbell and Fiske (1959). A principal components analysis with varimax rotation, was performed on the correlations for each school. In each analysis, two clear factors emerged: one risk factor with heavy loadings on the eight RTOOE measures, and one aptitude-achievement dimension with heavy loadings on the four legitimate measures. It appears that R' and R_z measured essentially the same trait throughout the four tests, and that the RTOOE trait was distinguishable from the dimension corresponding to the legitimate scores. Therefore, while there was some RTOOE behavior specific to each particular test, there was a large general factor that appeared in all four tests. The finding of this general factor confirms Swineford's results (1941).

It is important to note that the aptitude-achievement distinction, may not have been more than nominal. The general directions at the beginning of each test booklet attempted to distinguish between the aptitude (...A TEST WHICH WILL GIVE YOU A CHANCE TO SHOW WHAT YOU KNOW AND HOW WELL YOU THINK) and the achievement examinations (...A TEST OF SOME OF THE KNOWLEDGE YOU HAVE GAINED DURING YOUR SCHOOL YEARS). However, the dis-

Table 3.1

School Means on Risk (R', R_s) and Legitimate Measures (L)

		Males Schools						Females Schools						Total Schools					
		S_1	S_2	S_3	S_4	S_5	S_6	S_1	S_2	S_3	S_4	S_5	S_6	S_1	S_2	S_3	S_4	S_5	S_6
R'	Apt-L*	.649	.391	.495	.295	.473	.620	.709	.620	.611	.426	.496	.628	.678	.511	.562	.366	.485	.624
	Apt-M**	.636	.426	.538	.336	.488	.610	.691	.629	.541	.459	.504	.613	.663	.533	.540	.402	.496	.612
	Ach-L†	.709	.607	.712	.514	.658	.723	.800	.784	.721	.627	.684	.741	.754	.700	.718	.575	.672	.733
	Ach-M††	.509	.385	.470	.211	.429	.518	.620	.549	.505	.330	.471	.491	.563	.471	.491	.275	.450	.504
R_s	Apt-L	.832	.685	.756	.575	.631	.800	.885	.815	.827	.688	.702	.814	.858	.744	.797	.636	.668	.808
	Apt-M	.873	.715	.847	.725	.753	.831	.912	.880	.864	.768	.714	.843	.892	.802	.857	.748	.733	.837
	Ach-L	.863	.769	.838	.728	.740	.848	.901	.876	.872	.778	.793	.889	.882	.825	.858	.755	.767	.870
	Ach-M	.880	.731	.858	.649	.771	.803	.940	.852	.901	.747	.764	.825	.910	.794	.883	.702	.768	.814
L	Apt-L	15.40	14.14	17.88	14.64	9.12	14.84	17.27	16.15	21.14	14.99	10.59	17.45	16.32	15.20	19.78	14.83	9.88	16.24
	Apt-M	7.57	7.43	10.03	6.04	3.35	6.77	5.75	5.34	8.35	4.63	3.25	5.59	6.67	6.33	9.05	5.27	3.30	6.14
	Ach-L	17.94	15.89	18.32	15.27	13.47	14.52	18.57	18.16	20.71	15.92	15.33	16.73	18.25	17.09	19.72	15.63	14.42	15.70
	Ach-M	10.84	9.14	11.05	8.15	4.87	8.58	9.83	9.06	9.15	5.82	3.47	8.64	10.34	9.10	9.94	6.89	4.15	8.61
n =		47	46	40	56	48	60	46	51	56	66	51	69	93	97	96	122	99	129

*Aptitude-Language
**Aptitude-Mathematics
†Achievement-Language
††Achievement-Mathematics

Table 3.2

Correlations and Reliabilities for Risk (R', R_s) and Legitimate Measures (L): Total Group

			Apt-L*			Apt-M**			Ach-L†			Ach-M††		
		School	R'	L	R_s	R'	L	R_s	R'	L	R_s	R'	L	R_s
Aptitude-Language	R'	S_1	.93	-.42	.84	.85	-.41	.80	.75	-.41	.57	.79	-.34	.58
		S_2	.93	-.16	.82	.76	-.34	.64	.81	-.37	.61	.83	-.38	.53
		S_3	.91	-.16	.81	.77	-.33	.55	.71	-.30	.45	.67	-.34	.28
		S_4	.86	.05	.81	.72	-.12	.66	.67	-.30	.60	.53	-.13	.38
		S_5	.94	-.26	.84	.68	-.29	.60	.71	-.44	.69	.66	-.39	.62
		S_6	.95	-.36	.85	.77	-.33	.64	.76	-.44	.45	.73	-.33	.31
	L	S_1		.77	-.08	-.34	.48	-.33	-.39	.74	-.10	-.36	.47	-.12
		S_2		.83	.06	-.32	.44	-.00	-.08	.69	-.03	-.12	.51	.13
		S_3		.80	.02	-.17	.46	.10	-.12	.69	-.03	-.10	.57	.28
		S_4		.72	.20	-.03	.30	.08	-.01	.63	.18	-.13	.33	.19
		S_5		.61	-.31	-.30	.13	-.16	-.13	.48	-.31	-.37	.26	-.38
		S_6		.85	-.31	-.39	.47	-.30	-.32	.73	.09	-.35	.50	.10
	R_s	S_1			.89	.80	-.33	.75	.74	-.30	.68	.68	-.31	.61
		S_2			.94	.63	-.36	.55	.75	-.06	.74	.73	-.11	.56
		S_3			.83	.71	-.29	.60	.66	-.14	.49	.66	-.31	.37
		S_4			.88	.68	-.04	.74	.64	-.04	.66	.59	-.03	.57
		S_5			.91	.68	-.19	.63	.65	-.35	.72	.67	-.30	.62
		S_6			.91	.77	-.33	.69	.76	-.32	.44	.68	-.37	.36
Aptitude-Mathematics	R'	S_1				.91	-.43	.81	.76	-.33	.57	.76	-.37	.64
		S_2				.87	-.30	.77	.71	-.32	.49	.83	-.30	.54
		S_3				.90	-.36	.70	.68	-.31	.47	.79	-.32	.43
		S_4				.74	-.03	.80	.64	-.30	.61	.65	-.13	.55
		S_5				.84	-.16	.78	.63	-.34	.67	.74	-.32	.68
		S_6				.89	-.36	.82	.75	-.16	.50	.84	-.34	.41
	L	S_1					.54	-.40	-.50	.49	-.36	-.46	.53	-.30
		S_2					.31	-.33	-.33	.54	-.21	-.32	.55	-.11
		S_3					.05	[illegible]	-.38	.49	-.30	-.33	.74	.10
		S_4					.47	[illegible]	-.06	.27	-.06	-.05	.41	.05
		S_5					.13	-.14	-.14	.31	-.18	-.09	.42	-.09
		S_6					.50	-.34	-.31	.50	-.01	-.34	.52	-.07
	R_s	S_1						.89	.65	-.31	.55	.67	-.33	.69
		S_2						.82	.62	-.11	.41	.65	-.15	.60
		S_3						.70	.51	-.10	.45	.56	.31	.64
		S_4						.79	.62	-.11	.64	.60	-.03	.67
		S_5						.86	.58	-.34	.60	.56	-.16	.66
		S_6						.86	.69	-.18	.48	.69	-.17	.52
Achievement-Language	R'	S_1							.89	-.41	.67	.88	-.39	.65
		S_2							.88	-.11	.69	.81	-.35	.56
		S_3							.88	-.35	.62	.82	-.39	.31
		S_4							.78	-.13	.67	.62	-.34	.48
		S_5							.85	-.40	.79	.78	-.33	.75
		S_6							.91	-.33	.66	.77	-.33	.46
	L	S_1								.86	-.33	-.43	.62	-.24
		S_2								.89	-.06	-.18	.54	.04
		S_3								.83	-.33	-.30	.55	.04
		S_4								.70	-.18	-.18	.33	.06
		S_5								.76	-.46	-.38	.46	-.34
		S_6								.86	-.03	-.41	.59	.05
	R_s	S_1									.57	.64	-.16	.72
		S_2									.66	.61	-.14	.45
		S_3									.47	.57	-.38	.32
		S_4									.65	.50	-.11	.49
		S_5									.80	.73	-.33	.74
		S_6									.63	.49	.06	.38
Achievement-Mathematics	R'	S_1										.94	-.45	.69
		S_2										.94	-.21	.67
		S_3										.92	-.37	.51
		S_4										.86	-.33	.57
		S_5										.91	-.30	.81
		S_6										.94	-.39	.57
	L	S_1											.64	-.18
		S_2											.52	-.04
		S_3											.66	.15
		S_4											.45	-.13
		S_5											.68	-.18
		S_6											.58	.09
	R_s	S_1												.80
		S_2												.92
		S_3												.83
		S_4												.83
		S_5												.89
		S_6												.89

*Aptitude-Language
**Aptitude-Mathematics
†Achievement-Language
††Achievement-Mathematics

tinction between the two types of tests is not always obvious. In Table 3.1, it is seen that in every case, the aptitude examination in mathematics or language was more difficult than the corresponding achievement examination (Recall that there were 40 legitimate items on the language aptitude test, and 30 legitimate items on each of the other tests.) Therefore, the attempt to have less familiar items on the aptitude tests appears to have been successful, and it is clear that the aptitude-achievement dimension is confounded with difficulty. Indeed, some may argue that the aptitude and achievement tests differ in difficulty only, and not in any aptitude-achievement classification. This writer would not take issue with such people. There is no doubt, however, that the four tests did differ with respect to content (language versus mathematics) and in difficulty, and that there was a large general factor throughout these tests.

Also of some interest was the tendency, across all six schools, for the legitimate measures to correlate negatively with the RTOOE measures. In other words, high propensities in RTOOE tended to be associated with low legitimate scores; low propensities in RTOOE tended to be associated with high legitimate scores. These negative correlations were not, in general, of great magnitude. However, previous studies of the relation between RTOOE and legitimate measures have either found no correlation (Swineford, 1941), or slightly positive correlations (Slakter, 1967). A possible explanation for this discrepancy might be due to the lack of testwiseness at the 8th grade level, and the resultant inability to profit by taking risks. Another possible explanation is that some of the students, knowing that the test scores would not enter into grades, etc., answered all of the questions without attempting to select the correct answer. The latter explanation was investigated by examining several of the scatter diagrams after the elimination of all *S*s who were assigned RTOOE scores of 1 (i.e., high risk takers). Since the correlations still appeared negative, the conjecture that the negativeness was due to disinterested examinees was considered less plausible. The conjecture concerning the lack of testwiseness at the 8th grade level will be investigated in future research.

In another interesting side issue, a schools by sex multivariate factorial analysis of variance was performed on the R' vector; i.e., the set of dependent RTOOE measures was composed of R' on language aptitude, R' on mathematics aptitude, R' on language achievement, and R' on mathematics achievement. Since the cell frequencies were unequal, an exact least squares analysis was utilized (Bock, 1963). The computer program used for the calculations was prepared by Finn (1967). Using the .05 level, it was found that sex differences (with the effects of schools eliminated) were significant, school differences (with the effects of sex eliminated) were significant, but that the interaction (with both main effects eliminated) was not significant.

It is important to note that Swineford (1941) found males to be higher on RTOOE than females, while the present data indicate just the opposite; i.e., females were higher on RTOOE than males. This discrepancy in results may be explained by the fact that Swineford's *S*s were in 9th grade, while the *S*s of the present study were just entering 8th grade. Since very little is known of the relation between RTOOE and age, it is possible that the difference in findings is merely a function of the age of the *S*s. On the other hand, it should be remembered that in the Swineford study (1941, p 439), approximately 25% of the *S*s were eliminated. Some of these *S*s were eliminated because they neglected to ask for extra credit; i.e., were low on RTOOE. Since twice as many males were eliminated as females, there is reason to suspect that the mean RTOOE score for males was spuriously high.

From an inspection of Table 3.1, it is seen that the English-speaking Canadian School (S_4) was the lowest in RTOOE, with the bilingual Canadian school (S_5) also somewhat low in RTOOE. One might speculate that this result was due to differences in test sophistication, since the Canadian schools have been less exposed to objective examinations than their United States counterparts. In order to determine if the sex or school differences in RTOOE might be accounted for by the differences in legitimate score, a schools by sex multivariate factorial analysis of covariance was performed on the R' vector, with the legitimate scores on the four tests used as the covariates. As in the multivariate analysis of variance, the cell frequencies were unequal, and an exact least squares analysis was used. Also, as in the multivariate analysis of variance, significant differences were found for sex (with schools eliminated), for schools (with sex eliminated), but not for interaction (with both main effects eliminated). With the use of covariance analysis, however, the bilingual Canadian school appeared to be almost as low in RTOOE as the English-speaking Canadian school.

Finally, judging from both Tables 3.1 and 3.2, it appears that the New York State school whose testing time was shortened by five minutes (S_6), was relatively unaffected by the decrease in time.

CONCLUSIONS AND IMPLICATIONS

The major question of interest dealt with the generality of RTOOE across different types of tests. For various kinds of schools, it was found that a large, general RTOOE factor was present in the tests. In other words, *S*s who tended to be high (low) in RTOOE on a mathematics examination, tended to be high (low) in RTOOE on a language examination; *S*s who tended to be high (low) in RTOOE on aptitude (or difficult) examinations, tended to be high (low) in RTOOE on achievement (or less difficult examinations.

Therefore, there is some evidence that the specific type of testing situation was not too important in the measurement of RTOOE. If such is the case, the particular type of testing situation used for the measurement of RTOOE, might well be decided on the basis of convenience. For example, the language tests were composed largely of vocabulary items of the synonym-antonym type. This type of test is comparatively easy to construct, and can be administered in a relatively short time. Therefore, it seems reasonable to use these synonym-antonym type tests to measure RTOOE, rather than (say) an appropriate mathematics test, which might take considerably more time to construct, and more time to administer.

CHAPTER IV: EFFECT OF RTOOE ON TEST SCORE

INTRODUCTION

Quereshi (1960) has shown that personality factors such as risk taking, play an important part in mental test scores. An earlier study by Votaw (1936) demonstrated that students with submissive personalities tend to be penalized by do-not-guess directions because, when compared to ascendent students, they omitted more items to which they knew the answer. A similar study by Sherriffs and Boomer (1954) found essentially the same results for maladjusted or high anxiety students; i.e., the high anxious students when compared to the low anxious students, omitted more items to which they knew the answer. A more recent investigation (Slakter, submitted) discovered that students who displayed low RTOOE on the Concept Mastery Test, tend to be penalized on their test score.

Two recent experiments investigating the effect of RTOOE on test score report somewhat conflicting findings. Hammerton (1965) found that *S*s instructed to answer all questions, scored higher on vocabulary tests than *S*s instructed to leave out questions when not confident of the answer. On the other hand, Taylor (1966) studied three different instruction groups (do not answer unless certain; do as well as you can; answer all questions), and found no differences in test score due to instructions. These contradictory results by Taylor might be due to the tendency for partial information to be less of a factor in mathematics examinations than in other types of examinations; e.g., vocabulary examinations. However, the contradictory findings may also be explained by reference to a table presenting the number of omitted and unfinished items for each of the three experimental groups (Taylor, 1966, p 4). From an inspection of the table, it appears that an invalid statistical test was applied to the data to demonstrate that there were differences in RTOOE among the three groups. Furthermore, when the average number of omitted and unfinished items for each of the three groups was calculated from the data in the table, it appeared that there was little difference in RTOOE among the groups. The Taylor results, therefore, may be misleading because of the failure to effectively manipulate RTOOE with the directions.

The present experiment, in order to more thoroughly study the effect of RTOOE on test score, consisted of two parts. The first part was concerned with conventional directions; the second with the Coombs type directions (Coombs, 1953; Coombs, Milholland, & Womer, 1956); i.e., the examinee is instructed to select those alternatives that are incorrect, with resultant losses and gains in test points depending on whether or not the correct alternative is selected, and how many alternatives are selected. Specifically, the following RTOOE strategies were examined:

1. For conventional do-not-guess directions, which of the following strategies results in a higher average test score?
    - a. Answer all items that you think will help you to obtain as high a score as possible. (usual risk)
    - b. Answer all items. (high risk)
2. For Coombs type directions, which of the following strategies result in the highest average test score? (assume 4 alternatives for each item.)
    - a. Always select exactly one alternative. (low risk)
    - b. Always select exactly two alternatives. (medium risk)
    - c. Always select exactly three alternatives. (high risk)
    - d. Always select those and only those alternatives that you are sure are incorrect. (usual risk)

Ss

*S*s were 11th grade students in four suburban or rural schools in western New York State. In two of the schools, the entire class (377 *S*s, 131 *S*s) took part in the study; in one of the schools, an intact group of 159 participated. In the largest 11th grade class, a random sample of 150 was selected from the total group of 468. With the natural attrition due to absences, etc., the actual number participating was somewhat lower in each school.

PROCEDURE

In each school, *S*s were assigned at random to one of the six treatment groups. All *S*s then received the same test items;

the specific directions to be followed by the *S* depended upon the particular group to which the *S* had been assigned. For copies of the directions, see the appendix E. The test consisted of 100 vocabulary items, each with four alternatives. Copies of the test items are provided in the appendix F.

The tests were scored to provide each *S* with a "corrected" vocabulary score and a RTOOE score. For the conventional groups, the vocabulary score was found by subtracting one-third of the number of incorrect responses, from the number of correct responses; the RTOOE score was calculated by the use of Ziller's index (R_z). For the Coombs groups, the vocabulary score was found by subtracting three times the number of correct alternatives selected, from the number of incorrect alternatives selected. The RTOOE score for the Coombs groups was calculated by using the following modified Ziller index:

$$R'_z = \frac{4(\#\text{ correct alternatives selected})}{4(\#\text{ correct alternatives selected}) + \#\text{ incorrect alternatives omitted}} \; .$$

The testing time was approximately 50 minutes, with the "usual risk" group under the Coombs directions requiring the most time because their directions were the most complicated. Adequate time was given for all *S*s to finish.

RESULTS

The split-half reliabilities for the vocabulary score (odd versus even, "corrected" by the Spearman-Brown formula), for each of the two conventional treatment groups in each of the four schools, are presented in Table 4.1a. Values in parentheses are the sample sizes. In similar fashion, the vocabulary test reliabilities for the Coombs treatment groups, are provided in Table 4.1b. In each table, the rows correspond to the schools, and the columns to the treatment groups. The column descriptions correspond to those detailed on page 34; i.e., for Table 4.1b, the column titled "low" corresponds to the "always select exactly one alternative" group, the column titled "medium" corresponds to the "always select exactly two alternatives," etc.

Table 4.1

Split-half Reliabilities of Vocabulary Test
(Sample Sizes in Parentheses)

4.1a Conventional

RTOOE School	usual	high
S_1	.93 (56)	.95 (64)
S_2	.91 (28)	.90 (24)
S_3	.96 (17)	.96 (19)
S_4	.89 (20)	.93 (17)

4.1b Coombs

RTOOE School	low	medium	high	usual
S_1	.82 (55)	.88 (62)	.96 (60)	.94 (64)
S_2	.95 (29)	.84 (28)	.88 (28)	.85 (15)
S_3	.96 (18)	.88 (23)	.94 (17)	.98 (15)
S_4	.82 (18)	.72 (17)	.91 (20)	.94 (20)

Table 4.2a provides the mean RTOOE scores ($\overline{R}_z$) and the standard deviations (S_{R_z}) for each of the two conventional treatment groups, in each of the four schools. Values in parentheses are the standard deviations. For the Coombs directions, the corresponding mean RTOOE scores ($\overline{R}'_z$) and the standard deviations ($S_{R'_z}$), are presented in Table 4.2b.

Table 4.2

Mean RTOOE Score (Standard Deviation in Parentheses)

4.2a Conventional

RTOOE School	usual	high
S_1	.74 (.2216)	1.00 (.0072)
S_2	.77 (.2312)	1.00 (.0037)
S_3	.65 (.2679)	1.00 (.0017)
S_4	.72 (.2591)	1.00 (.0000)

4.2b Coombs

RTOOE School	low	medium	high	usual
S_1	.18 (.0855)	.48 (.1095)	1.00 (.0097)	.74 (.2603)
S_2	.18 (.1151)	.41 (.1206)	1.00 (.0046)	.72 (.2501)
S_3	.16 (.1329)	.46 (.1160)	1.00 (.0155)	.72 (.2653)
S_4	.16 (.0752)	.49 (.1028)	.97 (.1137)	.77 (.2355)

The mean vocabulary score for each of the two conventional treatment groups, in each of the four schools, is presented in Table 4.3; the mean vocabulary score for each of the Coombs treatment groups, in each of the four schools, is provided in Table 4.4.

Table 4.3

Mean Vocabulary Test Score for Conventional Groups

RTOOE School	usual	high
S_1	47.8	53.3
S_2	50.5	55.2
S_3	55.4	59.6
S_4	43.4	49.7

Table 4.4

Mean Vocabulary Test Score for Coombs Groups

RTOOE School	low	medium	high	usual
S_1	53.7	100.7	135.8	132.9
S_2	48.4	123.0	159.9	163.7
S_3	54.8	106.6	190.6	153.1
S_4	59.8	95.5	121.4	155.5

DISCUSSION

It is seen from an inspection of Table 4.1a, that the vocabulary test was generally reliable for the two types of conventional directions, in all four schools. Furthermore, Table 4.1b establishes essentially the same conclusion for the Coombs directions.

From an examination of Table 4.2a, it is clear that RTOOE was manipulated in the conventional directions phase. In the "high risk" group, if all *S*s followed directions, there would have been no omissions, and each *S* would therefore have received a RTOOE score of 1. The extent to which the *S*s in the "high risk" group followed directions is evidenced by the fact that the average R_z was 1.00 (to two decimal places), for each of the four schools. In addition, again because most of the *S*s followed directions in the "high risk" group, the standard deviation of the RTOOE measure was small in each of the four schools. In the "usual risk" conventional group, the average R_z ranged from .65 to .77, with standard deviations of approximately .25.

In the Coombs phase of the study, as in the conventional phase, all *S*s following directions in the "high risk" group, would have received a RTOOE score of 1. Once again, the extent to which the directions were followed is evidenced by the average R'_z values of 1.00 in three of the "high risk" groups, and .97 in the fourth (to two decimal places). In the "medium risk" group, the average R'_z ranged from .41 to .49, while in the "low risk" group, the average R'_z varied from .16 to .18. As expected, in these three groups with fixed RTOOE strategies, the standard deviations for the RTOOE scores were small. For the "usual risk" Coombs group, the average R'_z ranged from .72 to .77, with standard deviations of about .25. As in the conventional study, it is clear that RTOOE was manipulated. It is interesting to note, also, that the means and standard deviations for the modified Ziller in the Coombs "usual risk" group, were quite similar to the corresponding Ziller values in the conventional "usual risk" group.

One limitation of the Ziller and modified Ziller indices is that any *S* obtaining a perfect (vocabulary) test score, would achieve an undefined RTOOE score. Fortunately, in this study no *S* achieved a perfect test score, and therefore, no *S*s had to be eliminated from the study.

After establishing that the criterion vocabulary test was reliable, and that RTOOE was manipulated in both the conventional and Coombs studies, we now arrive at the major question of interest. Do these different RTOOE strategies affect average test score?

With the conventional directions, the results in Table 4.3 demonstrate that in each school, the "high risk" group averaged approximately five points higher than the "usual risk" group. A schools by RTOOE factorial analysis of variance was performed. Because the cell frequencies were unequal, an exact least squares analysis was utilized (Bock, 1963). The computer program used for the calculations was prepared by Finn (1967). The results indicated that RTOOE effects (with the effects of schools eliminated) were significant at the .05 level. The strength of relation between RTOOE and vocabulary test score was approximately .02. Neither schools (with the effects of RTOOE eliminated) nor the interaction (with both main effects eliminated) were significant at the .05 level.

In similar fashion, the results presented in Table 4.4 demonstrate that in each school, the "low risk" and the "medium risk" groups achieved lower scores than the "high risk" and "usual risk" groups. A schools by RTOOE factorial analysis of variance for unequal cells, indicated that the RTOOE effects (with the effects of schools eliminated), the school effects (with the effects of RTOOE eliminated), and the interaction (with both main effects eliminated) were each significant at the .05 level. However, the strength of relation between RTOOE and test score was approximately .35, while the strength of relation for schools and interaction were each less than .02. The mean vocabulary test scores for the "high risk" and "usual risk" groups were significantly higher (.05 level) than those of the "low risk" and "medium risk" groups. The failure of the "high risk" group to attain a significantly higher mean than the "usual risk" group (as in the conventional directions) was probably due to the fact that RTOOE with partial information under Coombs directions has a negative expected value, while the expected value of RTOOE with partial information under conventional directions, is positive.

In brief, the strength of relation between RTOOE and vocabulary test score was higher with the Coombs directions (.35) than with the conventional directions (.02), although both were significant at the .05 level. However, with the Coombs directions, RTOOE was better able to be manipulated over a wider range. One might speculate, therefore, that with better manipulation of RTOOE with the conventional directions, a greater strength of relation between RTOOE and test score would result.

CONCLUSIONS, IMPLICATIONS

In conclusion, whether the directions are of the conventional do-not-guess variety or of the Coombs type, the RTOOE strategies followed by examinees affect their vocabulary test scores. In general, it appears that examinees adopting a cautious RTOOE strategy will achieve lower test scores than if they had followed a less cautious RTOOE strategy.

The implications for the individual taking a test with the usual penalties for incorrect responses is clear: Answer all questions! Perhaps, however, the implication for the aptitude or achievement test constructor is even more compelling: If you include a penalty for incorrect responses, the test scores of the examinees will reflect their RTOOE strategies as well as their aptitudes or achievements.

CHAPTER V: SUMMARY

Frequently, objective examinations are administered with instructions cautioning the examinee of penalties for incorrect responses. Risk taking on objective examinations (RTOOE) occurs whenever the examinee responds to an item, without knowing the correct answer with certainty. This project was concerned with three aspects of RTOOE: (1) its measurement, (2) its generality, and (3) its effect on test score.

MEASUREMENT OF RTOOE

In the study of the measurement of RTOOE, a new index, based upon nonsense questions and conventional directions, was defined as the proportion of nonsense items attempted. It was pointed out that two previous studies had utilized nonsense questions with conventional directions. Although in each of these two studies, the central problem of interest was not the measurement of RTOOE, each study did provide data which could be used to demonstrate the reliability of the new index. Therefore, it was decided to investigate the concurrent validity of the new index by estimating its correlation with an index of RTOOE proposed by Ziller (1957).

The COLLEGE ENTRANCE VOCABULARY TEST (CEVT) was administered to entering freshmen at the State University of New York at Buffalo. Since the CEVT is composed of 20 nonsense items and 40 legitimate items, it was possible to obtain a measure of RTOOE by the new index, and a measure of RTOOE by the Ziller index.

For the 1,786 *S*s, the correlation between the two measures of RTOOE was estimated to be .76. The reliability found for the Ziller index indicated that about 70 percent of the variability in the Ziller measure was non-error. Therefore, from the square of the correlation between the two measures of RTOOE, we see that approximately $(.76)^2$ or 58 percent of the variability in the Ziller measure was predictable from the new index. In other words, much of the reliable variance of the Ziller measure was predictable from the new index. In addition, both measures of RTOOE were essentially uncorrelated with the legitimate score on the CEVT. Therefore, it was decided that the new index was a potentially valuable measure of RTOOE.

GENERALITY OF RTOOE

In considering RTOOE behavior, it is of interest to know whether the behavior is quite specific to the particular testing situation, or whether it is a more general trait. For this study, the question involved the generality of RTOOE across the following four types of tests: (1) language aptitude, (2) mathematics aptitude, (3) language achievement, and (4) mathematics achievement.

*S*s were students in the 8th grade from four western New York schools, and two Canadian schools. *S*s were administered the STANDARD EDUCATIONAL INTELLIGENCE TEST (SEIT), and the STANDARD EDUCATIONAL ACHIEVEMENT TEST (SEAT). Both SEIT and SEAT were composed of a language and mathematics section, and each section contained 10 nonsense items embedded in legitimate items. Therefore, each *S* received two RTOOE scores (nonsense index, and the Ziller index), and one legitimate score on each of the four examination sections. These 12 scores provided a 12 by 12 correlation matrix for each of the six schools.

An inspection of these 12 by 12 correlation matrices, with split-half reliabilities in the diagonals, indicated considerable evidence for convergent and discriminant validity for each of the six schools. In addition, principal components analysis with varimax rotations indicated one strong RTOOE factor, and one legitimate score factor, for each of the six schools. The results seemed to indicate, therefore, that RTOOE was general across the four different tests utilized for each of the six schools. Therefore, there is some evidence that the specific type of test situation used for the measurement of RTOOE is not too important. In other words, the test vehicle used for the measurement of RTOOE might be based on convenience; e.g., a vocabulary test, since vocabulary tests are comparatively easy to construct and quick to administer.

EFFECT OF RTOOE ON TEST SCORE

Previous studies have suggested that RTOOE affects test score. In order to investigate the effects of RTOOE on average test score, an experimental study was designed. In the experiment, RTOOE was manipulated, and the resultant effect on average test score noted. One part of the study involved conventional do-not-guess directions, while the other part was directed toward Coombs type directions. *S*s were 11th grade students in four suburban or rural schools in western New York State.

For the conventional directions, a schools by RTOOE factorial analysis of variance with unequal cell frequencies, indicated that RTOOE effects (with the effects of schools eliminated) were significant at the .05 level. The strength of relation between RTOOE and test score was approximately .02. For each school, the "high risk" group averaged five points higher than the "usual risk" group. Neither schools (with the effects of RTOOE eliminated) nor the interaction (with both main effects eliminated) were significant at the .05 level.

With the Coombs directions, a schools by RTOOE factorial analysis of variance with unequal cell frequencies, indicated that RTOOE (with the effects of schools eliminated), schools (with the effects of RTOOE eliminated) and the interaction (with both main effects eliminated) were each significant at the .05 level. However, the strength of relation between RTOOE and test score was approximately .35, while the strength of relation for schools and for interaction were each less than .02. The "high risk" and "usual risk" groups in each school had appreciably higher mean scores than their "low risk" and "median risk" counterparts.

The implication for the individual taking a test under conventional directions and the usual penalty for guessing is clear: Answer all questions! The implication for the test constructor is also clear: If you include a penalty for incorrect responses, the test scores of the examinees will reflect their RTOOE strategies as well as their aptitudes or achievements!

REFERENCES

Bock, R. D. Programming univariate and multivariate analysis of variance. *Technometrics*, 1963, *5*, 95-117.

Coombs, C. H. On the use of objective examinations. *Educational and Psychological Measurement*, 1953, *13*, 308-310.

Coombs, C. H., Milholland, J. E., and Womer, F. B. The assessment of partial knowledge. *Educational and Psychological Measurement*, 1956, *16*, 13-37.

Danielson, L. E. *Gambling proneness: Its measurement and expression in examination situations.* Unpublished doctoral dissertation, University of Michigan, 1956.

Finn, J. D. *Multivariance - univariate and multivariate analysis of variance and covariance: A Fortran IV program.* Unpublished manuscript, State University of New York at Buffalo, 1967.

Fruchter, B. Error scores as a measure of carefulness. *Journal of Educational Psychology*, 1950, *41*, 279-291.

Granich, L. A technique for experimenting on guessing in objective tests. *Journal of Educational Psychology*, 1931, *22*, 145-156.

Hammerton, M. The guessing correction in vocabulary tests. *British Journal of Educational Psychology*, 1965, *35*, 249-251.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Kogan, N., and Wallach, M. A. *Risk taking: A study in cognition and personality.* New York: Holt, Rinehart & Winston, 1964.

Ontario Department of Education. *Intermediate Division.* Ontario, Canada, 1951.

Quereshi, M. Y. Mental test performance as a function of payoff conditions, item difficulty, and degree of speeding. *Journal of Applied Psychology*, 1960, *44*, 65-77.

Sherriffs, A. C., and Boomer, D. S. Who is penalized by the penalty for guessing? *Journal of Educational Psychology*, 1954, *45*, 81-90.

Slakter, M. J. Risk taking on objective examinations. *American Educational Research Journal*, 1967, *4*, 31-43.

Slakter, M. J. The penalty for not guessing. Submitted, 1967.

Slakter, M. J., and Cramer, S. H. The relation between risk taking and vocational choice. In preparation.

Slovic, P. Convergent validation of risk taking measures. *Journal of Abnormal and Social Psychology*, 1962, *65*, 68-71.

Stone, L. A. Reliability of a utility of risk measure. *Psychological Reports*, 1962, *10*, 516.

Swineford, F. The measurement of a personality trait. *Journal of Educational Psychology*, 1938, *29*, 295-300.

Swineford, F. Analysis of a personality trait. *Journal of Educational Psychology*, 1941, *32*, 438-444.

Swineford, F., and Miller, P. M. Effects of directions regarding guessing on item statistics on a multiple choice vocabulary test. *Journal of Educational Psychology*, 1953, *44*, 129-139.

Taylor, P. H. A study of the effects of instructions in a multiple-choice mathematics test. *The British Journal of Educational Psychology*, 1966, *36*, 1-6.

Thorndike, E. L., and Lorge, I. *The teacher's word book of 30,000 words*. New York: Columbia University Press, 1944.

Thurstone, L. L. The perceptual factor. *Psychometrika*, 1938, *3*, 1-17.

Torrance, E. P., and Ziller, R. C. Risk and life experience: Development of a scale for measuring risk taking tendencies. *USAF PTRC Research Report*, 1957, No. 57-23.

University of the State of New York, State Education Department. *Mathematics 7-8-9*. Albany, New York, 1962.

Varty, J. W. Guessing on examinations--is it worthwhile? *Educational Forum*, 1946, *10*, 205-212.

Votaw, D. F. The effect of do-not-guess directions upon the validity of true-false or multiple-choice tests. *Journal of Educational Psychology*, 1936, *27*, 698-703.

Wiley, L., and Trimble, O. C. The ordinary objective test as a possible criterion of certain personality traits. *School and Society*, 1936, *43*, 446-448.

Ziller, R. C. A measure of the gambling response-set in objective tests. *Psychometrika*, 1957a, *22*, 289-292.

Ziller, R. C. Vocational choice and utility for risk. *Journal of Counseling Psychology*, 1957b, *4*, 61-64.

APPENDIX A

STANDARD EDUCATIONAL INTELLIGENCE ACHIEVEMENT TEST

GRADES 7 - 8 - 9

DIRECTIONS FOR ADMINISTERING

note: instructions which are to be read aloud to students are printed in large type. instructions printed in small type are intended only for examiner.

1. when the students are assembled in the testing room and seated, say:

THE TESTING PERIOD HAS BEGUN. THERE SHOULD BE NO TALKING AMONG YOU UNTIL AFTER YOU HAVE BEEN DISMISSED. DO NOT OPEN YOUR BOOKLET OR TURN IT OVER UNTIL YOU ARE TOLD TO DO SO. MAKE SURE THAT YOU HAVE AN ANSWER SHEET, SCRATCH PAPER, AND A PENCIL.

2. have students fill in the necessary identification on their answer sheets; name, date, school, etc. in addition, have each student enter a 1 in the first space of the "ID" section if male, and a 2 if female. also have the student enter a 1 in the second space to indicate the SEIT. then have all students enter the code number for the school into the third space. then say:

TURN TO THE GENERAL DIRECTIONS ON THE FRONT COVER. READ THESE DIRECTIONS SILENTLY WHILE I READ THEM ALOUD. (now read the general directions with the group) NOW TURN TO THE NEXT PAGE, AND READ THE DIRECTIONS SILENTLY WHILE I READ THEM ALOUD. (now read the directions with the group)

3. at the end of the directions, answer any legitimate questions. stay within the meaning, and as far as possible, the vocabulary of the printed directions. then say:

WHEN I SAY "BEGIN", TURN TO THE NEXT PAGE AND START WORKING. READY? BEGIN!

4. examiners should move quietly about the room to see that each student is marking his answers correctly.

5. after approximately 1 hour of time has elapsed, check to see if most of the group is finished. do not let them spend more than 1 hour and 10 minutes on the first test. say:

STOP. PLEASE PUT YOUR PENCILS DOWN.

6. collect answer sheets, test booklets, for the intelligence test; pass out booklets containing answer sheets for the achievement test. allow students to relax at their seats but caution them not to talk. when all of the achievement booklets are passed out to the students, have students fill in the necessary identification on their answer sheets; name, date, school, etc. in addition, have each student enter a 1 in the first space of the "ID" section if male, and a 2 if female. also have the student enter a 2 in the second space to indicate the SEAT. then have all students enter the code number for the school into the third space. then say:

TURN TO THE GENERAL DIRECTIONS ON THE FRONT COVER. READ THESE DIRECTIONS SILENTLY WHILE I READ THEM ALOUD. (now read the directions with the group). THE DIRECTIONS FOR THIS TEST ARE THE SAME AS FOR THE PREVIOUS TEST; I.E., CORRECT ANSWERS RECEIVE 1 POINT, INCORRECT ANSWERS LOSE 1/3 OF A POINT, AND OMISSIONS NEITHER RECEIVE NOR LOSE POINTS. ANY QUESTIONS? WHEN I SAY "BEGIN", START WORKING. READY" BEGIN!

7. again the examiner should move quietly about the room to make sure that each student is marking his answers correctly.

8. after 2 hours and 15 minutes has elapsed, check to see if most of the group is finished. if most are finished or if we are rushed for time, say:

STOP. PLEASE PUT YOUR PENCILS DOWN.

9. collect answer sheets, test booklets, pencils, and discard scratch paper. write down for the record a description of any unexpected variation from the normal testing procedure that may have occured.

APPENDIX B

STANDARD EDUCATIONAL INTELLIGENCE TEST
GRADES 7-8-9

FORM T
LANGUAGE-ARITHMETIC

GENERAL DIRECTIONS

THIS BOOKLET CONTAINS A TEST WHICH WILL GIVE YOU A CHANCE TO SHOW WHAT YOU KNOW AND HOW WELL YOU THINK. MAKE SURE YOU UNDERSTAND THE TEST DIRECTIONS BEFORE YOU BEGIN WORKING. YOU MAY ASK QUESTIONS ABOUT ANY PART OF THE DIRECTIONS YOU DO NOT UNDERSTAND.

DIRECTIONS

EACH OF THE QUESTIONS OR INCOMPLETE STATEMENTS IN THIS TEST IS FOLLOWED BY FOUR SUGGESTED ANSWERS. YOU ARE TO DECIDE WHICH ONE OF THESE FOUR ANSWERS YOU SHOULD CHOOSE. YOU MUST MARK ALL OF YOUR ANSWERS ON THE SEPARATE ANSWER SHEET YOU HAVE BEEN GIVEN; THIS TEST BOOKLET SHOULD NOT BE MARKED IN ANY WAY. YOU MUST MARK YOUR ANSWER BY BLACKENING THE SPACE HAVING THE SAME NUMBER AS THE ANSWER YOU HAVE CHOSEN.

YOU MAY FIND SOME OF THE QUESTIONS VERY EASY AND SOME OF THEM RATHER DIFFICULT. *YOU WILL RECEIVE ONE POINT FOR EACH CORRECT ANSWER. FOR EACH INCORRECT ANSWER, YOU WILL LOSE 1/3 OF A POINT. IF YOU OMIT A QUESTION, YOU WILL NEITHER RECEIVE NOR LOSE A POINT.*

FOR EXAMPLE:

200. WHICH ONE OF THE FOLLOWING IS AN ANIMAL?

1) TABLE 2) ERASER 3) DOG 4) PAPER

SINCE A DOG IS AN ANIMAL, YOU SHOULD CHOOSE THE ANSWER NUMBERED 3) ON YOUR ANSWER SHEET. YOU WOULD FIRST FIND THE ROW OF SPACES NUMBERED THE SAME AS THE QUESTION--IN THE EXAMPLE ABOVE, IT IS 200. THEN YOU WOULD BLACKEN THE SPACE IN THIS ROW WHICH HAS THE SAME ANSWER AS THE ANSWER YOU HAVE CHOSEN. IN OTHER WORDS, IT WOULD APPEAR ON THE ANSWER SHEET AS:

200. 1 ═ 2 ═ 3 ▬ 4 ═

FOR THE CORRECT ANSWER ABOVE, YOU WOULD RECEIVE ONE POINT (+1). ON THE OTHER HAND, IF YOU HAD CHOSEN ANY ANSWER OTHER THAN THE CORRECT ANSWER, YOU WOULD HAVE LOST 1/3 OF A POINT (-1/3). FINALLY, IF YOU HAD DECIDED TO OMIT THE QUESTION BECAUSE YOU DID NOT KNOW THE CORRECT ANSWER, YOU WOULD HAVE NEITHER RECEIVED NOR LOST POINTS (0). NOTE, THEREFORE, THAT AN INCORRECT RESPONSE IS PENALIZED MORE THAN AN OMISSION.

BE SURE TO MAKE YOUR ANSWER MARKS HEAVY AND BLACK. MARK ONLY ONE ANSWER FOR EACH QUESTION. IF YOU CHANGE YOUR MIND ABOUT AN ANSWER, ERASE THE FIRST MARK COMPLETELY.

THE EXAMPLE ABOVE HAS BEEN GIVEN TO YOU SO THAT YOU WILL KNOW HOW TO MARK YOUR ANSWER SHEETS AND HOW THE ANSWERS ARE SCORED. THE QUESTIONS ON THE INSIDE OF THE TEST ARE NOT JUST LIKE THE EXAMPLE; BUT EACH ONE DOES HAVE FOUR CHOICES, AND YOU MUST CHOOSE YOUR ANSWER FROM AMONG THEM. REMEMBER, *YOU WILL RECEIVE ONE POINT FOR EACH CORRECT ANSWER. FOR EACH INCORRECT ANSWER, YOU WILL LOSE 1/3 OF A POINT. IF YOU OMIT A QUESTION, YOU WILL NEITHER RECEIVE NOR LOSE A POINT. ALWAYS TRY TO OBTAIN AS HIGH A SCORE AS POSSIBLE.*

ARE THERE ANY QUESTIONS?

LANGUAGE

Select the word which is most like the capitalized word.

1. DISCONTENT:
   1) pleasure 2) disorder 3) baffled 4) dissatisfaction
2. TEDIOUS:
   1) tiresome 2) temporary 3) thrilling 4) terrifying
3. VIOLENT:
   1) purple 2) scruff 3) rough 4) weak
4. HOMATE
   1) safeguard 2) reside 3) burden 4) fragment
5. LURK:
   1) bilk 2) pluck 3) perk 4) skulk
6. JEALOUSY:
   1) sympathy 2) envy 3) interest 4) anger
7. IRATE:
   1) logical 2) distinctive 3) cost 4) wrathful
8. VANITY:
   1) slowness 2) conceit 3) stupidity 4) piquancy
9. DEMJURIN:
   1) glad 2) studious 3) treacherous 4) responsible
10. PUZZLE:
    1) expel 2) dictate 3) hasten 4) confuse
11. PERFIMALY:
    1) illness 2) wretched 3) irregularity 4) acrimony
12. EAGER:
    1) aloof 2) rich 3) young 4) anxious
13. BICTEL:
    1) dilute 2) harden 3) decorate 4) improve
14. MYCARIUM:
    1) fiance´ 2) building 3) growth 4) vehicle
15. SOLEMN:
    1) serious 2) joyful 3) special 4) important
16. LINESTUS:
    1) saving 2) curved 3) cumbersome 4) reclining

GO ON TO NEXT PAGE

17. CHASTISE:
1) punish 2) pure 3) run 4) crude

18. IRREGULAR:
1) unusual 2) difficult 3) invisible 4) uneven

19. DEFORMED:
1) defoliate 2) defunct 3) distorted 4) cheated

20. FUMBLE:
1) rave 2) smoke 3) disinfect 4) grope

21. PROSARE
1) burden 2) aid 3) delude 4) labor

22. IMPETUOUS:
1) inadequate 2) begging 3) insular 4) headlong

23. CONSOLE:
1) bathe 2) comfort 3) amuse 4) feed

24. SPECIFIC:
1) suitable 2) essential 3) particular 4) elementary

25. DUBIOUS:
1) genuine 2) questionable 3) double 4) little

26. PALIDATE:
1) frighten 2) throw 3) appear 4) eat

27. SECEDE:
1) withdraw 2) insure 3) plant 4) kill

28. BILIOUS:
1) sad 2) double 3) greedy 4) bitter

29. SECRET:
1) frank 2) clandestine 3) devoted 4) profane

30. ECCENTRIC:
1) strict 2) original 3) clever 4) peculiar

31. EVICT:
1) eject 2) emulate 3) edify 4) evoke

32. COMPLEX:
1) rounded 2) satisfied 3) involved 4) necessary

33. GNARLED:
1) hollow 2) rotted 3) twisted 4) splintered

34. GUARDIAN:
1) holy 2) custodian 3) soldier 4) shelter

GO ON TO NEXT PAGE

35. HOBBLE:
1) blend 2) climb 3) limp 4) peek

36. MENACE:
1) threaten 2) capture 3) surround 4) destroy

37. SAUNTER:
1) dash 2) stagger 3) stroll 4) skip

38. EQUITABLE:
1) reasonable 2) despondent 3) biased 4) exempt

39. INFORMAL:
1) casual 2) advisory 3) inflated 4) structured

40. MAGNETIC:
1) generous 2) rugged 3) delicious 4) attractive

41. QUIFFENT:
1) incompetent 2) surprising 3) drink 4) separation

42. FABRICATE:
1) grease 2) gawk 3) build 4) barter

43. EXTRAVAGANTLY:
1) willingly 2) wastefully 3) cautiously 4) secretly

44. MATE:
1) insure 2) slander 3) join 4) muffle

45. SEVER:
1) tangle 2) splice 3) cut 4) tighten

46. AMIDDITY:
1) oddity 2) peaceful 3) essence 4) doubt

47. QUERY:
1) question 2) warning 3) message 4) answer

48. LENIENT:
1) severe 2) merciful 3) fair 4) respected

49. JUBILANT:
1) mischievous 2) brilliant 3) guilty 4) rejoicing

50. ACADEMIC:
1) recent 2) theoretical 3) popular 4) pious

GO ON TO NEXT PAGE

ARITHMETIC

Select the best answer.

51. Which of the following is a plane figure?

1) pyramid
2) trapezoid
3) sphere
4) cylinder

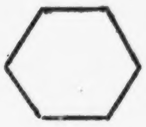
52. Which group of angles could be used to construct a triangle?

1) 100°, 60°, 40°
2) 90°, 50°, 40°
3) 60°, 40°, 20°
4) 50°, 30°, 20°

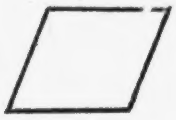
53. Of the geometric figures at the right, which is a pentagon?

1) A
2) B
3) C
4) D

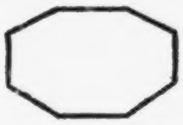
A



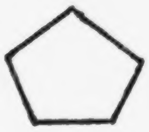
B



C



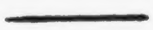
D

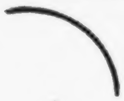


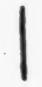
54. Which of the following is an example of a differentiated fraction?

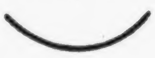
1) 3/5
2) .54
3) 2 1/3
4) a/b

55. An example of a vertical line would be:

1) 

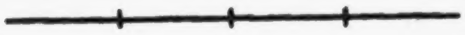
2) 

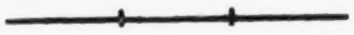
3) 

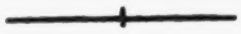
4) 

56. Which pair of line segments are in the ratio of 2 to 1?

1) A and C
2) B and C
3) A and B
4) none of these

A 

B 

C 

GO ON TO NEXT PAGE

57. The largest number below is:

1) .25
2) 3/5
3) 2/3
4) .8

58. If a circle is drawn with a compass spread 3 inches, what part of the circle would be 3 inches in length?

1) arc
2) circumference
3) radius
4) diameter

59. An interpolator is used for measuring:

1) distance
2) volume
3) area
4) weight

60. If 1 out of 10 people prefer item x, what percent of 60 people should prefer item x?

1) 1
2) 6
3) 10
4) 16 2/3

61. How many cents are there in d dimes?

1) 10 + d
2) d/10
3) 10/d
4) 10d

62. Which of the following values is not included in the range represented by 5 ± 0.5?

1) 4.5
2) 4.8
3) 5.0
4) 5.6

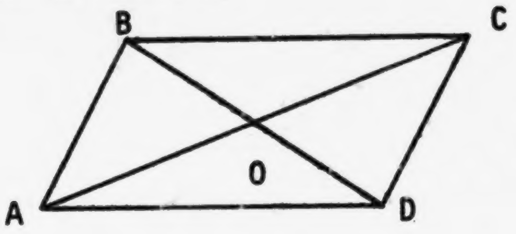
63. A measurement in rods may be converted to a measurement in:

1) centimeters
2) grams
3) liters
4) square yards

GO ON TO NEXT PAGE

64. Which of the angles in the figure at the right is obtuse?

1) ABC
2) COD
3) BOD
4) DAC



65. The parallactive law of arithmetic is best represented by the statement:

1) $3 \cdot (4+7) = 3 \cdot 4 + 3 \cdot 7$
2) $5+7 = 7+5$
3) $(9+5) - 3 = 9 + (5-3)$
4) $6 \cdot 1/6 = 1$

66. Which of the following fractions would be a repeating decimal?

1) 3/4
2) 5/6
3) 3/5
4) 7/8

67. The formula that would be used to determine the area of a biozoid is:

1) $A = \pi/d$
2) $A = h/3 \cdot b^2$
3) $A = 2\pi/r$
4) $A = s^3/2$

68. Which two numbers are both factors of 10?

1) 9, 1
2) 2, 5
3) 5, 4
4) 8, 2

69. Henry answered 6 problems incorrectly on a certain test, but did 80% of them correctly. How many problems were on the test?

1) 30
2) 86
3) 74
4) 68

70. A hostess needs enough ice cream to serve 54 persons. How many quarts should she order if one quart will serve 6 persons?

1) 324
2) 9
3) 60
4) 48

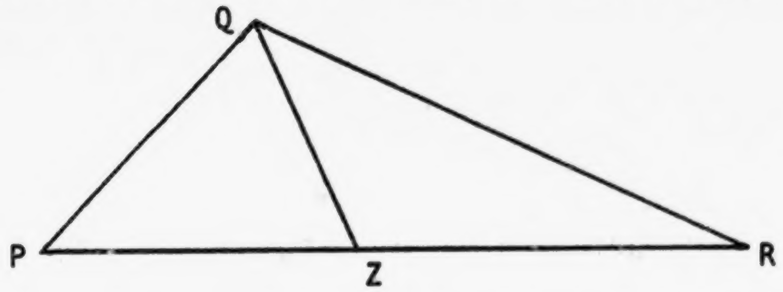
GO ON TO NEXT PAGE

71. Which of the following would be an example of an infinite set?

1) the set of even numbers greater than 10
2) the set of all people now alive on earth
3) {1,2,3,4,---,99}
4) {A,B,C,D,E,---,Z}

72. In triangle PQR at the right, PZ = ZR. The mode of this triangle would be:

1) line PR
2) line QZ
3) point Q
4) angle PQR



73. The cost of an article is reduced 20%. What fraction of the original price is the new price?

1) 19/20
2) 4/5
3) 1/5
4) 1/20

74. Which equation can be solved by adding 9 to both sides or members?

1) $x + 9 = 17$
2) $x/9 = 17$
3) $x-9 = 17$
4) $9x = 17$

75. What is the reciprocal of 25?

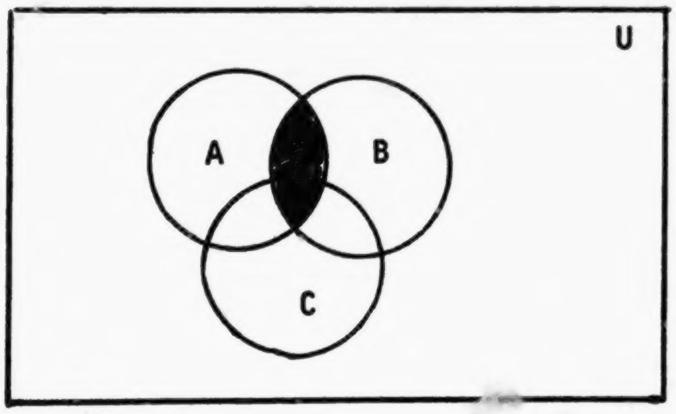
1) -25
2) 1/5
3) 1/25
4) 25 has no reciprocal

76. The symbol Γ means:

1) add
2) radian
3) centolis
4) propulate

77. The shaded area on the diagram at the right represents what relationship among the sets A, B, and C?

1) $A \cup B$
2) $A \cap B$
3) $B \cup C$
4) $B \cap C$



GO ON TO NEXT PAGE

78. The number of times an integer is used in a multiplication problem is referred to as a (an):
   1) coefficient
   2) exponent
   3) term
   4) equation

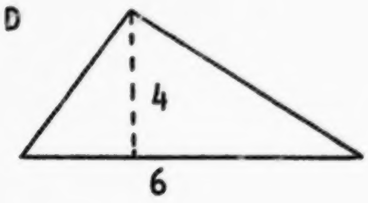
79. 4 + 6 ÷ 2 - 3 equals:
   1) 4
   2) 2
   3) 5
   4) 7/2

80. The truncated root of 8 is:
   1) 2
   2) 4
   3) $4\sqrt{2}$
   4) $\sqrt{2}$

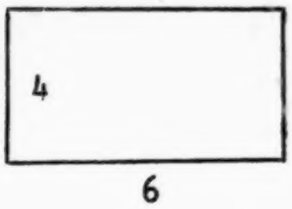
81. The starting five players for the championship basketball team had heights of 6 ft. 2 in., 6 ft. 1 in., 6 ft. 8 in., 5 ft. 11 in., and 6 ft. What is the team's average height?
   1) 6 feet
   2) 6 feet 1 inch
   3) 6 feet 2 inches
   4) 6 feet 3 inches

82. Which two figures have the same area?
   1) B and D
   2) C and D
   3) A and C
   4) A and D

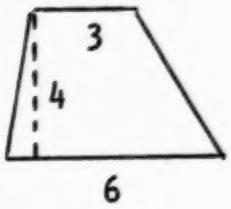
A



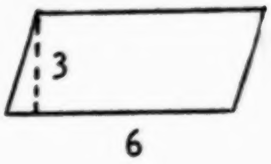
B



C



D



83. Which of the following terms is related to the measure of an invariant segre?
   1) mile
   2) diameter
   3) congruent
   4) circumference

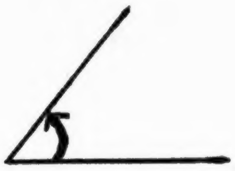
84. At 6 A.M. the temperature was -7°, at noon it was $+2^{\circ}$. How many degrees did the temperature change during the morning?
   1) 9
   2) 5
   3) 2
   4) 7

GO ON TO NEXT PAGE

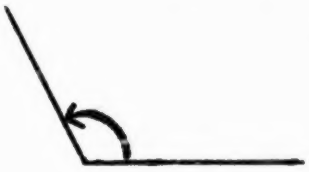
85. 8.82 ÷ .6 = 14.7 . What would the new answer be if the decimal point was omitted from the dividend?
   1) .147
   2) 147
   3) 1.47
   4) 1470

86. Which is an example of a disjoint angle?
   1) A
   2) B
   3) C
   4) D

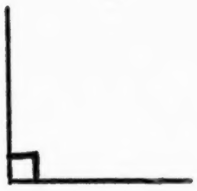
A



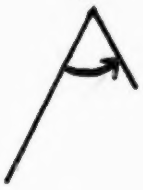
B



C



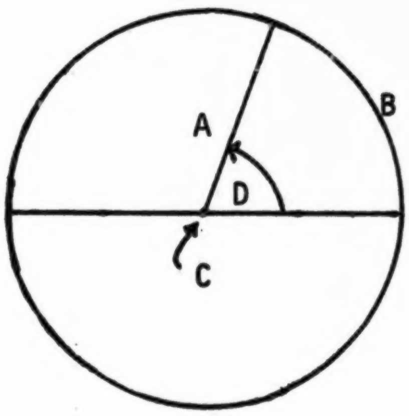
D



87. The minuend is 5, the subtrahend is 14, the difference is:
   1) 19
   2) 70
   3) -9
   4) -19

88. What part of 32 is a quarter of 48?
   1) 1/4
   2) 12
   3) 3/8
   4) 3/4

89. What part of the figure at the right is the frustrum?
   1) A
   2) B
   3) C
   4) D



90. Which of the following is a trinomial:
   1) 1+2+3
   2) 1/3
   3) 5•5•5
   4) $\sqrt[3]{27}$

APPENDIX C

STANDARD EDUCATIONAL ACHIEVEMENT TEST
GRADES 7-8-9

FORM K
LANGUAGE-ARITHMETIC

GENERAL DIRECTIONS

THIS BOOKLET CONTAINS A TEST OF SOME OF THE KNOWLEDGE YOU HAVE GAINED DURING YOUR SCHOOL YEARS. MAKE SURE YOU UNDERSTAND THE TEST DIRECTIONS BEFORE YOU BEGIN WORKING. YOU MAY ASK QUESTIONS ABOUT ANY PART OF THE DIRECTIONS YOU DO NOT UNDERSTAND.

DIRECTIONS

EACH OF THE QUESTIONS OR INCOMPLETE STATEMENTS IN THIS TEST IS FOLLOWED BY FOUR SUGGESTED ANSWERS. YOU ARE TO DECIDE WHICH ONE OF THESE FOUR ANSWERS YOU SHOULD CHOOSE. YOU MUST MARK ALL OF YOUR ANSWERS ON THE SEPARATE ANSWER SHEET YOU HAVE BEEN GIVEN; THIS TEST BOOKLET SHOULD NOT BE MARKED IN ANY WAY. YOU MUST MARK YOUR ANSWER BY BLACKENING THE SPACE HAVING THE SAME NUMBER AS THE ANSWER YOU HAVE CHOSEN.

YOU MAY FIND SOME OF THE QUESTIONS VERY EASY AND SOME OF THEM RATHER DIFFICULT. *YOU WILL RECEIVE ONE POINT FOR EACH CORRECT ANSWER. FOR EACH INCORRECT ANSWER, YOU WILL LOSE 1/3 OF A POINT. IF YOU OMIT A QUESTION, YOU WILL NEITHER RECEIVE NOR LOSE A POINT.*

FOR EXAMPLE:

200. WHICH ONE OF THE FOLLOWING IS AN ANIMAL?

1) TABLE 2) ERASER 3) DOG 4) PAPER

SINCE A DOG IS AN ANIMAL, YOU SHOULD CHOOSE THE ANSWER NUMBERED 3) ON YOUR ANSWER SHEET. YOU WOULD FIRST FIND THE ROW OF SPACES NUMBERED THE SAME AS THE QUESTION--IN THE EXAMPLE ABOVE, IT IS 200. THEN YOU WOULD BLACKEN THE SPACE IN THIS ROW WHICH HAS THE SAME ANSWER AS THE ANSWER YOU HAVE CHOSEN. IN OTHER WORDS, IT WOULD APPEAR ON THE ANSWER SHEET AS:

200. 1 ═ 2 ═ 3 ▬ 4 ═

FOR THE CORRECT ANSWER ABOVE, YOU WOULD RECEIVE ONE POINT (+1). ON THE OTHER HAND, IF YOU HAD CHOSEN ANY ANSWER OTHER THAN THE CORRECT ANSWER, YOU WOULD HAVE LOST 1/3 OF A POINT (-1/3). FINALLY, IF YOU HAD DECIDED TO OMIT THE QUESTION BECAUSE YOU DID NOT KNOW THE CORRECT ANSWER, YOU WOULD HAVE NEITHER RECEIVED NOR LOST POINTS (0). NOTE, THEREFORE, THAT AN INCORRECT RESPONSE IS PENALIZED MORE THAN AN OMISSION.

BE SURE TO MAKE YOUR ANSWER MARKS HEAVY AND BLACK. MARK ONLY ONE ANSWER FOR EACH QUESTION. IF YOU CHANGE YOUR MIND ABOUT AN ANSWER, ERASE THE FIRST MARK COMPLETELY.

THE EXAMPLE ABOVE HAS BEEN GIVEN TO YOU SO THAT YOU WILL KNOW HOW TO MARK YOUR ANSWER SHEETS AND HOW THE ANSWERS ARE SCORED. THE QUESTIONS ON THE INSIDE OF THE TEST ARE NOT JUST LIKE THE EXAMPLE; BUT EACH ONE DOES HAVE FOUR CHOICES, AND YOU MUST CHOOSE YOUR ANSWER FROM AMONG THEM. REMEMBER, *YOU WILL RECEIVE ONE POINT FOR EACH CORRECT ANSWER. FOR EACH INCORRECT ANSWER, YOU WILL LOSE 1/3 OF A POINT. IF YOU OMIT A QUESTION, YOU WILL NEITHER RECEIVE NOR LOSE A POINT. ALWAYS TRY TO OBTAIN AS HIGH A SCORE AS POSSIBLE.*

ARE THERE ANY QUESTIONS?

LANGUAGE

A. Identify the underlined portions of the following sentences as one of the choices numbered 1 through 4 below. Mark your answer sheet to correspond to the numbered choices.

1. subject
2. compound preposition
3. direct object
4. declarative predicate

1. Shakespeare wrote *Macbeth* and *Henry V*.
2. Mrs. Smith left a pie on the table.
3. You missed by a mile.
4. Does this furnace burn oil?
5. Mary certainly will help her sister.
6. The girl in the blue dress is going shopping.
7. Jim and Pat saw the sky-divers.
8. He traded the gun for a knife.
9. Mr. Jones cancelled his two o'clock appointment.
10. Get a car.
11. Have you ever seen a spotted zebra?
12. They went for a ride in the car.
13. For his birthday, Max received some tickets for the championship game.
14. My little brother lost his green tricycle.
15. Wells and Black tried to set a new record in the Grand Prix.
16. The athletes enjoyed their visit to Japan.

GO ON TO NEXT PAGE

B. Read the following selections. Following each are some questions based upon that selection.

> The seasons gradually rolled on. The trees budded, blossomed, and bore fruit. The swallows came, built nests, reared their young, and winged on.

17. Of the following, the main idea of the above paragraph is best expressed as

    1) all life is a struggle.
    2) changes are seen in the passage of the seasons.
    3) trees bud and birds fly.
    4) changes are almost always for the better.

18. The logical expression of selected conjunctions in the above paragraph would most likely lead to the statement that

    1) all life is a struggle.
    2) changes are seen in the passage of the seasons.
    3) trees bud and birds fly.
    4) changes are almost always for the better.

19. The above paragraph is composed of three sentences.

    1) all are simple sentences
    2) all are compound sentences
    3) first sentence is simple; second and third sentences are compound
    4) first and third sentences are simple; second sentence is compound

20. Rearrangement of the comma modifiers in the sentences of the above paragraph would allow you to say that

    1) all are simple sentences.
    2) all are compound sentences.
    3) first sentence is simple; second and third sentences are compound.
    4) first and third sentences are simple; second sentence is compound.

*** *** *** *** ***

GO ON TO NEXT PAGE

A very common idea is that certain crops must be planted in the dark of the moon, while others must be planted in the full moon. Scientists have not been able to find that the moon has anything to do with crops. It makes no difference to plants whether the moon is shining or not.

21. The main idea of the above paragraph is that

    1) the moon has no effect on plants.
    2) plants on the moon do not shine.
    3) certain crops can be planted at night.
    4) farmers plant crops.

22. Which of the following underlined words is an example of a redefined preposition in the above paragraph?

    1) in the dark of the moon (line 1)
    2) in the full moon (line 2)
    3) to find (line 3)
    4) to plants (line 4)

GO ON TO NEXT PAGE

C. Select the word which is most like the capitalized word.

23. BARTER:
    1) lift   2) trade   3) buy   4) sell

24. CRAVE:
    1) want   2) need   3) hole   4) scare

25. INAUDIBLE:
    1) unheard   2) uneaten   3) unemployed   4) undone

26. REDAPT:
    1) know   2) turn   3) drink   4) call

27. COMMAND:
    1) give   2) receive   3) order   4) teach

28. SOAR:
    1) grunt   2) fly   3) hurt   4) bleed

29. GENEROUS:
    1) talkative   2) liberal   3) friendly   4) fast

30. SHRILL:
    1) scoff   2) stiff   3) creek   4) high

31. FEUD:
    1) fight   2) old   3) definite   4) burn

32. EXNECATE:
    1) dissolve   2) behead   3) revert   4) retreat

33. IMMENSE:
    1) wave   2) leaf   3) bump   4) huge

34. FALTER:
    1) harness   2) stagger   3) drop   4) gash

35. BICENE:
    1) forge   2) decline   3) man   4) obscure

36. IMAGE:
    1) picture   2) boat   3) book   4) candle

37. ACCUMULATE:
    1) puzzle   2) forget   3) astonish   4) collect

38. HIDDEN:
    1) obituary   2) odd   3) obscure   4) odious

39. INGORIAN:
    1) subtle   2) definite   3) sad   4) mad

40. WRATH:
    1) anger   2) flower   3) destroy   4) lawn

GO ON TO NEXT PAGE

ARITHMETIC

Select the best answer.

41. In decimal notation the number MCXXI is:

    1) 1121
    2) 10121
    3) 5121
    4) 10521

42. The number seven hundred sixty-five million, three hundred nine thousand, eight hundred forty-seven in digital notation would be:

    1) 7,653,908,047
    2) 765,309,847
    3) 76,539,847
    4) 765,390,847

43. What number would result if 297,568 is rounded to the nearest thousand?

    1) 290,000
    2) 297,000
    3) 298,000
    4) 300,000

44. 8/7 is an example of what kind of fraction?

    1) proper
    2) mixed
    3) integral
    4) improper

45. The sum of $8\frac{3}{10} + 4\frac{5}{6}$ equals:

    1) $13\frac{2}{15}$
    2) $12\frac{8}{16}$
    3) $12\frac{8}{10}$
    4) none of these

46. In the problem 5279 - 2304 = 2975, the subtrahend would be:

    1) 2304
    2) 5279
    3) 2975
    4) none of these

47. To divide by a fraction, you invert and:

    1) divide
    2) add
    3) multiply
    4) subtract

GO ON TO NEXT PAGE

48. The term certemplate is usually associated with which of the following arithmetic problems?

1) square root
2) logarithms
3) multiplication
4) division

49. $6.45 + 3.1 \div 8$ equals:

1) .684
2) 17.55
3) 9.63
4) none of these

50. Which of the following do the symbols { } represent?

1) polygons
2) sets
3) congruence
4) square roots

51. 5 is what percent of 40?

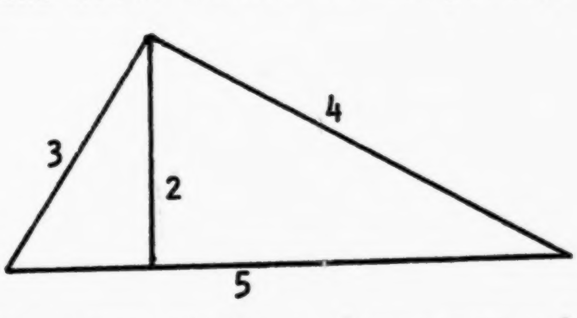
1) 8
2) 5
3) 3.2
4) $12\frac{1}{2}$

52. A protometer would be used to measure:

1) an angle of elevation
2) arc length
3) directrix of an ellipse
4) the speed of an automobile

53. Which of the following computations would result in the perceptil of the figure at the right?

1) 2×5
2) $2 + 3 + 4$
3) $2 + 5$
4) $2 \times (3 + 4)$

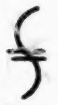


54. The chart below shows the batting record of four players on a baseball team during two games. If batting average equals the quotient of the number of hits and the number of times at bat, who had the highest average?

1) Dick
2) Sid
3) Bill
4) Tom

PLAYER	TIMES AT BAT	NO. OF HITS
Dick	7	3
Sid	11	5
Bill	8	5
Tom	9	4

GO ON TO NEXT PAGE

55. The symbol  is used to represent a (an):

1) angle
2) cubit
3) frustrum
4) centroid

56. *Charm* magazine reports a circulation of 660,000. There were actually 655,838 magazines sold. To what accuracy is the number above rounded?

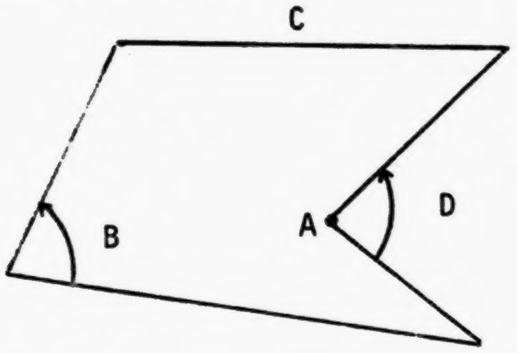
1) nearest hundred
2) nearest thousand
3) nearest ten thousand
4) nearest hundred thousand

57. A basketball team won 75% of their games. If they won 45 games, how many games did they play?

1) 34
2) 60
3) 30
4) 120

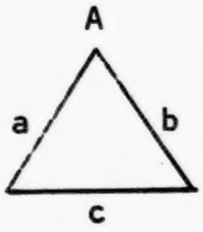
58. What part of the figure at the right is the centrosis?

1) A
2) B
3) C
4) D

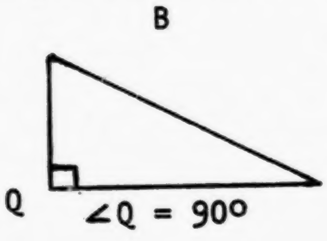


59. An example of a scalene triangle would be:

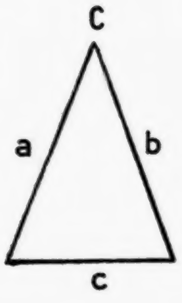
1) A
2) B
3) C
4) D



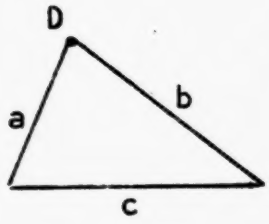
a=b=c



∠Q = 90°



a = b



a ≠ b ≠ c

60. How many sides does a contralateral have:

1) 5
2) 7
3) 9
4) 2

GO ON TO NEXT PAGE

61. Overhead is to profit as legrity is to:

1) commission
2) interest
3) discount
4) margin

62. 3/4 of a ton equals how many pounds?

1) 900
2) 3960
3) 1500
4) 1320

63. Mr. Stapleton bought a lawn mower listed at $80, receiving a 15% discount. If he paid 2% sales tax, what is his total bill?

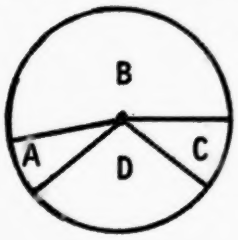
1) $69.36
2) $67.00
3) $69.60
4) none of these

64. Which of the following is a necessary property of a line segment?

1) has no midpoint
2) must have end points
3) must be curved
4) has no end points

65. The graph at the right shows how Johnny spends his earnings. Which portion of the graph would show the 33 1/3% Johnny spends for entertainment?

1) A
2) B
3) C
4) D



66. Orange juice was on sale at 4 cans for 39 cents. If the regular price is 11 cents per can, how many cents did Mrs. Smith save by buying 12 cans at the sale price?

1) 28
2) 15
3) 3
4) 12

67. Which of the following terms is usually not associated with angles?

1) ray
2) vertex
3) obtuse
4) width

GO ON TO NEXT PAGE

68. Which of the objects below has a capacity closest to one gallon?

1) tea cup
2) one-pound coffee can
3) man's shoebox
4) bath tub

69. Our number system is based on 10 whereas the Babylonian system was based on 60. Which system of measure below best resembles the Babylonian system?

1) English weight
2) English length
3) U.S. money
4) time

70. $L = \frac{3h}{4r}$ is the formula used for determining:

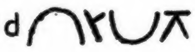
1) area of a monolateral
2) perimeter of a cycloid
3) surface area of a trisepher
4) area of a fortran

71. The term minex is related to which of the following?

1) arithmetic
2) measurement
3) geometry
4) graphs

72. 1/6 is the additive inverse of:

1) 6
2) 0
3) -1/6
4) 5/6

73. What number system would  be a member of?

1) Arabic
2) Egyptian
3) Icelandic
4) Babylonian

74. A cube has how many surfaces?

1) 4
2) 8
3) 6
4) 12

75. Determine the value of w in the proportion 7: 10 = 42: w?

1) 6
2) 60
3) 40
4) 70

GO ON TO NEXT PAGE

76. A student did 6 out of 10 problems correctly. What is the ratio of the number right to the number wrong?

1) 3/5
2) 3/2
3) 2/3
4) 5/3

77. If each side of a rectangle is doubled, the area is:

1) doubled
2) multiplied by 4
3) halved
4) equal to the perimeter

78. The maximum number of obtuse angles a triangle may have is:

1) 1
2) 2
3) 3
4) 0

79. Geometric constructions involve the use of which pair of tools?

1) ruler, protractor
2) computer, dividers
3) angulator, vernier caliper
4) straight edge, compass

80. Centimeter is to inch as kilometer is to:

1) foot
2) yard
3) mile
4) acre

APPENDIX D

STANDARD EDUCATIONAL INTELLIGENCE TEST

10 - 11 - 12

DIRECTIONS FOR ADMINISTERING

note: instructions which are to be read aloud to students are printed in large type. instructions printed in small type are intended only for the examiner.

1. as students enter the room, ask them to select a desk with a test paper having the same number as that on their name card (not necessary if only 1 group in room).
2. when the students are assembled in the examination room and seated, say:

THE TESTING PERIOD HAS BEGUN. THERE SHOULD BE NO TALKING AMONG YOU UNTIL AFTER YOU HAVE BEEN DISMISSED. DO NOT OPEN YOUR BOOKLET OR TURN IT OVER UNTIL
(not necessary if only 1 group)
YOU ARE TOLD TO DO SO. (MAKE SURE THAT THE NUMBER ON THE TEST BOOKLET MATCHES THE NUMBER ON YOUR NAME CARD.) PLEASE CHECK TO MAKE SURE THAT YOU HAVE AN ANSWER SHEET, AND A PENCIL.

3. have students fill in the necessary identification on their answer sheets; name, date, school, form (e.g., K-3, etc.). in addition have each student enter a 1 in the first space of the "ID" section if male, and a 2 if female. also have the student enter his form number in the second space in the "ID" section (e.g., 1, 2, 3, 4, 5, 6). then have all students enter code number for school into third space. then say:

TURN TO THE GENERAL DIRECTIONS ON THE FRONT COVER. READ THESE DIRECTIONS SILENTLY WHILE I READ THEM ALOUD. (now read the general directions with the group). NOW TURN TO THE NEXT PAGE, AND READ THE DIRECTIONS SILENTLY WHILE I READ THEM ALOUD. (now read the directions with the group, being careful to read only directions common to all treatments, and asking the students to silently read those that differ with each treatment.)

4. at the end of the directions, answer any legitimate questions. stay within the meaning, and, as far as possible, the vocabulary of the printed directions. then say:

   THERE ARE 100 QUESTIONS ON THE TEST. WHEN I SAY "BEGIN", TURN TO THE NEXT PAGE AND START WORKING. READY? BEGIN!

5. examiner should move quietly about the room to see that each student is following the appropriate directions, and that he is marking his answers correctly.

6. when there is 10 minutes of testing time remaining, ask the group to plan on finishing in 10 minutes. when the 10 minute period has expired (assuming most have finished or we are rushed for time) say:

   STOP. PLEASE PUT YOUR PENCILS DOWN.

   (otherwise allow students to continue until most are finished, and record the amount of time taken)

7. collect answer sheets, test booklets, and pencils. write down for the record a description of any unexpected variation from the normal testing procedure that may have occurred.

APPENDIX E

STANDARD EDUCATIONAL INTELLIGENCE TEST

GRADES 10-11-12

FORM K-1

LANGUAGE

GENERAL DIRECTIONS

This booklet contains a test which will give you a chance to show what you know and how well you think. Make sure you understand the test directions before you begin working. You may ask questions about any part of the directions you do not understand.

DIRECTIONS

Each question number has a word to the right of it, and four other words immediately below. You are to choose the word from the set of four words, which has the same meaning, or most nearly the same meaning, as the word to the right of the question number. You must mark all of your answers on the separate answer sheet you have been given; this test booklet should not be marked in any way. You must mark your answer by blackening the space having the same number as the answer you have chosen.

You may find some of the questions very easy and some of them rather difficult. You will receive one point for each correct answer. For each incorrect answer, you will lose 1/3 of a point. If you omit a question, you will neither receive not lose a point.

For example:

200. TOSS:

1) subtract 2) drink 3) throw 4) leave

Since the word "throw" has most nearly the same meaning as toss, you should choose the answer numbered 3) on your answer sheet. You would first find the row of spaces numbered the same as the question--in the example above, it is 200. Then you would blacken the space in this row which has the same number as the answer you have chosen. In other words, it would appear on the answer sheet as:

200. 1 ═ 2 ═ 3 ▬ 4 ═

For the correct answer above, you would receive one point (+1). On the other hand, if you had chosen any answer other than the correct answer, you would have lost 1/3 of a point (-1/3). Finally, if you had decided to omit the question because you did not know the correct answer, you would have neither received not lost points (0). Note, therefore, that an incorrect answer is penalized more than an omission.

Mark only one answer for each question. If you change your mind about an answer, erase the first mark completely.

Remember, you will receive one point for each correct answer, for each incorrect answer, you will lose 1/3 of a point. If you omit a question, you will neither receive nor lose a point. Do not spend too much time on any one question. Always try to obtain as high a score as possible.

Are there any questions?

STANDARD EDUCATIONAL INTELLIGENCE TEST

GRADES 10-11-12

FORM K-2

LANGUAGE

GENERAL DIRECTIONS

This booklet contains a test which will give you a chance to show what you know and how well you think. Make sure you understand the test directions before you begin working. You may ask questions about any part of the directions you do not understand.

DIRECTIONS

Each question number has a word to the right of it, and four other words immediately below. You are to choose the word from the set of four words, which has the same meaning, or most nearly the same meaning, as the word to the right of the question number. You must mark all of your answers on the separate answer sheet you have been given; this test booklet should not be marked in any way. You must mark your answer by blackening the space having the same number as the answer you have chosen.

For example:

200. TOSS:

1) subtract   2) drink   3) throw   4) leave

Since the word "throw" has most nearly the same meaning as toss, you should choose the answer numbered 3) on your answer sheet. You would first find the row of spaces numbered the same as the question--in the example above, it is 200. Then you would blacken the space in this row which has the same number as the answer you have chosen. In other words, it would appear on the answer sheet as:

200. 1═   2═   3▬   4═

Mark only one answer for each question. If you change your mind about an answer, erase the first mark completely.

You may find some of the questions very easy and some of them rather difficult. Answer every question, whether you are sure of the answer or not. You are not expected to be able to answer every question correctly. Do not spend too much time on any one question. Always do your best.

Are there any questions?

STANDARD EDUCATIONAL INTELLIGENCE TEST

GRADES 10-11-12

FORM K-3

LANGUAGE

GENERAL DIRECTIONS

This booklet contains a test which will give you a chance to show what you know and how well you think. Make sure you understand the test directions before you begin working. You may ask questions about any part of the directions you do not understand.

DIRECTIONS

Each question number has a word to the right of it, and four other words immediately below. One of the words, in the set of four words, has the same meaning, or most nearly the same meaning, as the word to the right of the question number. You are to choose one, and only one, word from the set of four words, that does *not* have the same, or most nearly the same, meaning as the word to the right of the question number. You must mark all of your answers on the separate answer sheet you have been given; this test booklet should not be marked in any way. You must mark your answers by blackening the space having the same number as the answer you have chosen.

For example:

200. TOSS:
1) subtract 2) drink 3) throw 4) leave

Since the word "subtract" does not have the same meaning as TOSS, you could choose the answer numbered 1) on your answer sheet. You would first find the row of spaces numbered the same as the question--in the example above, it is 200. Then you would blacken the space in this row which has the same number as the answer you have chosen. In other words, it would appear on the answer sheet as:

200. 1 ▬ 2 ═ 3 ═ 4 ═

Note also that either "drink" or "leave" would have been correct responses.

Mark one and only one choice for each question. If you change your mind about an answer, erase the first mark completely.

Remember, you are to select one word that does not have the same meaning as the word to the right of the question number. Answer every question, whether you are sure of the answer or not. You are not expected to be able to answer every question correctly. Do not spend too much time on any one question. Always do your best.

Are there any questions?

STANDARD EDUCATIONAL INTELLIGENCE TEST

GRADES 10-11-12

FORM K-4

LANGUAGE

GENERAL DIRECTIONS

This booklet contains a test which will give you a chance to show what you know and how well you think. Make sure you understand the test directions before you begin working. You may ask questions about any part of the directions you do not understand.

DIRECTIONS

Each question number has a word to the right of it, and four other words immediately below. One of the words, in the set of four words, has the same meaning, or most nearly the same meaning, as the word to the right of the question number. You are to choose two, and only two, words from the set of four words, that do *not* have the same meaning as the word to the right of the question number. You must mark all of your answers on the separate answer sheet you have been given; this test booklet should not be marked in any way. You must mark your answers by blackening the spaces having the same numbers as the answers you have chosen.

For example:

200. TOSS:

1) subtract 2) drink 3) throw 4) leave

Since the words "subtract" and "leave" do not have the same meaning as TOSS, you could select the choices numbered 1) and 4) on your answer sheet. You would first find the row of spaces numbered the same as the question--in the example above, it is 200. Then you would blacken the two spaces in this row which have the same numbers as the choices you have selected. In other words, it would appear as:

200. 1 ▬ 2 = 3 = 4 ▬

Note also that "subtract" and "drink", or "drink" and "leave" would have been correct responses.

Mark two and only two choices for each question. If you change your mind about an answer, erase the first mark completely.

Remember, you are to select two words that do not have the same meaning as the word to the right of the question number. Answer every question, whether you are sure of the answer or not. You are not expected to be able to answer every question correctly. Do not spend too much time on any one question. Always do your best.

Are there any questions?

STANDARD EDUCATIONAL INTELLIGENCE TEST

GRADES 10-11-12

FORM K-5

LANGUAGE

GENERAL DIRECTIONS

This booklet contains a test which will give you a chance to show what you know and how well you think. Make sure you understand the test directions before you begin working. You may ask questions about any part of the directions you do not understand.

DIRECTIONS

Each question number has a word to the right of it, and four other words immediately below. One of the words in the set of four words has the same meaning, or most nearly the same meaning, as the word to the right of the question number. You are to choose the three words in the set of four words, that do *not* have the same meaning as the word to the right of the question number. You must mark all of your answers on the separate answer sheet you have been given; this test booklet should not be marked in any way. You must mark your answers by blackening the spaces having the same numbers as the answers you have chosen.

For example:

200. TOSS:

1) subtract 2) drink 3) throw 4) leave

Since the words "subtract", "drink", and "leave" do not have the same meaning as TOSS, you would select the choices numbered 1), 2), and 4) on your answer sheet. You would first find the row of spaces numbered the same as the question--in the example above, it is 200. Then you would blacken the three spaces in this row which have the same numbers as the choices you have selected. In other words, it would appear as:

200. 1 ▬ 2 ▬ 3 = 4 ▬

Mark exactly three choices for each question. If you change your mind about an answer, erase the first mark completely.

Remember, you are to select the three words that do not have the same meaning as the word to the right of the question number. Answer every question, whether you are sure of the answer or not. You are not expected to be able to answer every question correctly. Do not spend too much time on any one question. Always do your best.

Are there any questions?

STANDARD EDUCATIONAL INTELLIGENCE TEST

GRADES 10-11-12

FORM K-6

LANGUAGE

GENERAL DIRECTIONS

This booklet contains a test which will give you a chance to show what you know and how well you think. Make sure you understand the test directions before you begin working. You may ask questions about any part of the directions you do not understand.

DIRECTIONS

Each question number has a word to the right of it, and four other words immediately below. One of the words in the set of four words has the same meaning or nearly the same meaning, as the word to the right of the question number. Let us call this word the like alternative, and the remaining three words will be called the unlike alternatives. You are to choose all those words and only those words that you are sure are *unlike* alternatives. Therefore for each question, you may select none, one, two, or three unlike alternatives, depending upon the amount of information that you have. One advantage of this method of selecting the unlike alternatives over the more conventional method of selecting the like alternative, is that you never have to guess. You choose all those words, and only those words, that you are sure are unlike alternatives.

For each unlike alternative that you choose, you will receive one point (+1). However, if you incorrectly select the like alternative as an unlike alternative, you will be penalized three points (-3).

You must mark all of your answers on the separate answer sheet you have been given; this test booklet should not be marked in any way. You must mark your answer(s) by blackening the space(s) having the same number(s) as the unlike alternative(s) you have selected.

For example:
200. TOSS:
1) subtract 2) drink 3) throw 4) leave

In this case, "subtract", "drink" and "leave" are unlike alternatives for "toss"; "throw" is a like alternative for "toss".

If you are sure that "subtract" is an unlike alternative, and you are not sure of the other unlike alternatives, then you would select the choice numbered 1) on your answer sheet. You would first find the row of spaces numbered the same as the question--in the example above, it is 200. Then you would blacken the space in this row which has the same number as the unlike alternative that you have selected. In other words, it would appear as:

200. 1▬ 2 ═ 3 ═ 4 ═

Since the word "subtract" is an unlike alternative, you would have received one point (+1) for blackening the space corresponding to 1). Furthermore, if you had selected as your single choice, either "drink" or "leave" you also would have received one point. On the other hand, had you selected as your single choice the like alternative "throw", you would have lost three points (-3).

In similar fashion, if you had selected two unlike alternatives, (e.g., "subtract", "drink") you would have received two points (+2). However, if you had selected two words, one an unlike alternative and the other the like alternative, (e.g., "subtract", "throw") then you would have received one point for the unlike alternative and lost three points for the like alternative, for a net loss of two points (-2) for the question.

Likewise, if you had selected the three unlike alternatives, (i.e., "subtract", "drink", "leave"), you would have received three points (+3). Of course, had you selected three choices, with the like alternative as one of the choices, (e.g., "subtract", "drink", "throw"), then you would have received two points for the unlike alternatives and lost three points for the like alternative, for a net loss of one point (-1) for the question. In short, you will receive one point for each unlike alternative you select, and you will lose three points for each like alternative you select.

If you change your mind about an answer, erase the first mark completely.

Remember, you are to select as many choices as you want as long as you are sure that you have not selected the like alternative. For each unlike alternative that you select, you will receive one point; for each like alternative you select, you will lose three points. Do not spend too much time on any one question. Always do your best.

Are there any questions?

APPENDIX F

1. VIOLENT:
1) purple 2) scruff 3) rough 4) weak

2. PUZZLE:
1) expel 2) dictate 3) hasten 4) confuse

3. SOLEMN:
1) serious 2) joyful 3) special 4) important

4. SPECIFIC:
1) suitable 2) essential 3) particular 4) elementary

5. SECRET:
1) frank 2) concealed 3) succinct 4) profane

6. INFORMAL:
1) casual 2) advisory 3) inflated 4) structured

7. MAGNETIC:
1) generous 2) rugged 3) delicious 4) attractive

8. KINDLE:
1) ignite 2) cut 3) know 4) knit

9. PETTY:
1) inferior 2) hard 3) pretty 4) capricious

10. SACCHARINE:
1) sacral 2) sweet 3) seasoned 4) granulose

11. WELD:
1) unite 2) wax 3) leap 4) scoop

12. DEFRAUD:
1) swindle 2) pledge 3) cultivate 4) detach

13. SECEDE:
1) withdraw 2) succeed 3) discriminate 4) attain

14. TEDIOUS:
1) tiresome 2) temporary 3) thrilling 4) terrifying

15. IRATE:
1) logical 2) distinctive 3) interested 4) wrathful

16. IRREGULAR:
1) abnormal 2) difficult 3) invisible 4) intensive

17. DEFORMED:
1) defoliated 2) defunct 3) distorted 4) cheated

18. FUMBLE:
1) rave 2) smoke 3) disinfect 4) grope

19. CONSOLE:
1) bathe 2) comfort 3) amuse 4) feed

20. EVICT:
1) eject 2) emulate 3) edify 4) evoke

21. COMPLEX:
1) rounded 2) satisfied 3) involved 4) necessary

22. FABRICATE:
1) grease 2) gawk 3) build 4) barter

23. EXTRAVAGANTLY:
1) willingly 2) wastefully 3) cautiously 4) effortlessly

24. LENIENT:
1) severe 2) merciful 3) fair 4) respected

25. BESTOW:
1) cover 2) hide 3) search 4) give

26. TWITCH:
1) scratch 2) whip 3) jerk 4) crunch

27. STEER:
1) rip 2) look 3) guide 4) strive

28. LIMBER:
1) flexible 2) limpid 3) passive 4) loathe

29. STRESS:
1) indicate 2) stare 3) strike 4) emphasize

30. ACCOMPANY:
1) praise 2) ascend 3) escort 4) accommodate

31. APPREHEND:
1) know 2) judge 3) appoint 4) arrest

32. SHABBY:
1) false 2) sincere 3) shapely 4) unkempt

33. BADGER:
1) annoy 2) hunt 3) escape 4) adorn

34. CONCUR:
1) confide 2) agree 3) run 4) sway

35. HALE:
1) pale 2) healthy 3) hedonic 4) spiritual

36. PLAUSIBLE:
1) convenient 2) believable 3) pliable 4) sufficient

37. DEVOUT:
1) determined 2) naive 3) involved 4) pious

38. CRUMBLE:
1) argue 2) decay 3) battle 4) cover

39. OFFICIAL:
1) mandated 2) sublimated 3) authorized 4) confiscated

40. POSTULATE:
1) resume 2) assume 3) mitigate 4) regulate

41. FACETIOUS:
1) fallacious 2) fantastic 3) mealy 4) witty

42. STYMIE:
1) filthy 2) reduce 3) gregarious 4) impede

43. GRATE:
1) grapple 2) induce 3) irritate 4) gratinate

44. TRIED:
1) tired 2) noble 3) forlorn 4) trusted

45. YEARN:
1) withdraw 2) primp 3) long 4) lean

46. PRUNE:
1) trim 2) destroy 3) activate 4) procreate

47. CONCILIATORY:
1) choleric 2) coarse 3) unfettered 4) peaceful

48. FACSIMILE:
1) smile 2) facility 3) fallacy 4) copy

49. HOSTILE:
1) gingery 2) clinical 3) witless 4) antagonistic

50. RUSTIC:
1) rural 2) red 3) rusty 4) poisonous

51. MAIM:
1) mutilate 2) condition 3) magnify 4) judge

52. REPUDIATE:
1) denote 2) renounce 3) duplicate 4) restore

53. PLIABLE:
1) plastic 2) plated 3) edible 4) passage

54. MEDITATE:
1) threaten 2) repair 3) intercede 4) ponder

55. SPONTANEOUS:
1) illustrious 2) courageous 3) instinctive 4) argumentative

56. RESPECTABLE:
1) honorable 2) solvent 3) repulsive 4) excitable

57. CONTORT:
1) warp 2) reform 3) endure 4) transport

58. CHASTE:
1) sociable 2) jubilant 3) virtuous 4) divine

59. PERSEVERE:
1) complicate 2) declaim 3) continue 4) convince

60. IRKSOME:
1) obstinate 2) complex 3) annoying 4) ingoing

61. SEDENTARY:
1) industrious 2) inferior 3) neat 4) placid

62. REITERATE:
1) replace 2) relate 3) displace 4) repeat

63. HARKEN:
1) sever 2) persecute 3) listen 4) observe

64. LURK:
1) work 2) pluck 3) perk 4) sneak

65. DUBIOUS:
1) genuine 2) vague 3) ductile 4) little

66. ECCENTRIC:
1) strict 2) original 3) clever 4) peculiar

67. GNARLED:
1) hollow 2) rotted 3) twisted 4) splintered

68. HOBBLE:
1) blend 2) climb 3) limp 4) peek

69. MENACE:
1) threaten 2) capture 3) surround 4) destroy

70. EQUITABLE:
1) reasonable 2) despondent 3) biased 4) exempt

71. CHIDE:
1) scold 2) scoff 3) scare 4) cheer

72. CUFF:
1) slap 2) kick 3) tap 4) butt

73. RUE:
1) regret 2) reform 3) remember 4) regress

74. REEL:
1) stagger 2) swagger 3) slip 4) catch

75. ADMONISH:
1) seclude 2) increase 3) control 4) reprove

76. OVERT:
1) dull 2) hasty 3) vicious 4) apparent

77. WHITTLE:
1) wonder 2) whisk 3) pare 4) signal

78. FOREIGN:
1) fractional 2) bespectacled 3) frugal 4) extrinsic

79. LAMENT:
1) lounge 2) limp 3) mourn 4) consider

80. GLIB:
1) gloomy 2) grand 3) fluent 4) leathery

81. STERN:
1) bigoted 2) silent 3) relentless 4) devious

82. PALATABLE:
1) studious 2) incidental 3) savory 4) peaceable

83. APPRECIABLE:
1) contemptible 2) perceptible 3) resentful 4) upright

84. LITERAL:
1) malleable 2) considerable 3) lateral 4) exact

85. DISSIDENT:
1) discourteous 2) affected 3) dejected 4) discordant

86. VIGILANT:
1) negligent 2) sublime 3) attentive 4) justifiable

87. CONDUCIVE:
1) intelligible 2) instrumental 3) concrete 4) conclusive

88. ABSOLVE:
1) acquit 2) absorb 3) foreclose 4) discard

89. CHASTISE:
1) punish 2) harden 3) decorate 4) dilute

90. IMPETUOUS:
1) inadequate 2) impossible 3) implicit 4) rash

91. CAUSTIC:
1) sarcastic 2) dry 3) causal 4) effective

92. PERT:
1) lively 2) sharp 3) proud 4) short

93. FOREBODE:
1) portend 2) inhabit 3) frighten 4) castigate

94. AQUILINE:
1) hooked 2) watery 3) acidic 4) majestic

95. TOUSLE:
1) rumble 2) crumble 3) rumple 4) fumble

96. LOLL:
1) droop 2) laud 3) lower 4) break

97. INVETERATE:
1) inverse 2) habitual 3) spineless 4) nonconforming

98. PRESAGE:
1) foretell 2) know 3) shorten 4) sweeten

99. MITIGATE:
1) translate 2) imitate 3) mock 4) moderate

100. VERBOSE:
1) redundant 2) authentic 3) proficient 4) unqualified